A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

The Measurement of Preconscious Perception in Dreams and Images: An Investigation of the Poetzl Phenomenon 285
Howard Shevrin and Lester Luborsky

Affective Response to Psychoanalytic Death Symbols 295
W. W. Meissner

The Specificity of Intolerance of Ambiguity Measures 300
Douglas T. Kenny and Ross Ginsberg

Perceptual Reactivity to Threat and Neutral Words Under Varying Experimental Conditions 305
Emory L. Cowen and Paul A. Obrist

Authoritarianism, Punitiveness, and Perceived Social Status 311
Alan H. Roberts and Richard Jessor

Conformity and Problem Solving 315
Charles Y. Nakamura

Continued on inside front cover

This is the Last Issue of Volume 56.
Title Page and Contents for the Volume
Appear Herein.

Volume 56 MAY 1958 Number 3

CONTENTS—*Continued*

A Restatement and Test of Schlosberg's Theory of Emotion With Two Kinds of Subjects From Greece 321
Harry C. Triandis and William W. Lambert

Psychological Test and Organismic Correlates of Interview Interaction Patterns 329
Ruth G. Matarazzo, Joseph D. Matarazzo, George Saslow, and Jeanne S. Phillips

Interrelationships Among Individual Difference Variables, Behavior in Psychotherapy, and Verbal Conditioning 339
Irwin G. Sarason

Interrelations Among Measures of Personality Adjustment in Nonclinical Populations 345
Fred E. Fiedler, John S. Dodge, Robert E. Jones, and Edwin B. Hutchins

Interacting Effects of Control, Motivation, Group Practice, and Problem Difficulty on Attempted Leadership 352
Bernard M. Bass, Margaret W. Pryer, Eugene L. Gaier, and Austin W. Flint

Effects of Motivating Instructions on Reaction Time in Cerebral Disease 359
Harold L. Blackburn

Autonomic Feedback: The Perception of Autonomic Activity 367
George Mandler, Jean M. Mandler, and Ellen T. Uviller

Intrusion of Associative Responses Into Schizophrenic Conceptual Performance 374
Loren J. Chapman

The Relationship Between Judgments of a Communicator's Position and Extent of Opinion Change 380
Walter Weiss

CASE REPORT

Was Lars Sexually Assaulted? A Study in the Reliability of Witnesses and of Experts 385
Arne Trankell

CRITIQUE AND NOTES

Accuracy of Self-Perception and Group-Perception as Related to Two Leadership Roles 396
Daphne E. Bugental and George F. J. Lehner

Behavior and Grammatical Style 398
Leonard W. Doob

Achievement Striving, Social Class, and Test Anxiety 401
Martin L. Hoffman, Spiro B. Mitsos, and Roland E. Protz

Volume 56 MAY, 1958 Number 3

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

DANIEL J. LEVINSON
Harvard Medical School

GARDNER LINDZEY
University of Minnesota

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Circulation Manager: HELEN ORR
Editorial Assistant: BARBARA B. CUMMINGS

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1958 by the American Psychological Association, Inc.

THE MEASUREMENT OF PRECONSCIOUS PERCEPTION IN DREAMS AND IMAGES: AN INVESTIGATION OF THE POETZL PHENOMENON[1]

HOWARD SHEVRIN AND LESTER LUBORSKY

Menninger Foundation

IN a pioneer laboratory investigation of dream imagery, Poetzl (16) discovered that his subjects dreamt about the originally unreported parts of a previously exposed picture. He flashed colored slides of everyday scenes at $\frac{1}{100}''$ and asked subjects to describe and draw what they saw. He then told them to record any dreams they had that night. When he asked his subjects on the following day to describe and draw their dreams, he observed that the dream imagery borrowed extensively from the unreported portions of the picture. Poetzl's subjects confirmed his observations by identifiying elements of the picture of which they were previously unaware as parts of their dream imagery.

Poetzl looked upon his findings as supporting the view that "images of indirect vision, errors of tachistoscopic vision in normal persons, visual dream-images and the visual mistakes of patients with agnosia agree with each other in many qualities regarding form and contents" (16). He attributed these similarities to a functional equivalence between brain damage to visual areas and the impoverishment of normal perception produced by tachistoscopic exposure or peripheral vision. In a brief exposure, involuntary eye movements from the point of fixation to other areas of the stimulus cannot be executed. The time limitation erects a "motoric barrier" that permits a clear perception of only part of the picture; the remainder of the picture registers in the visual apparatus but is not represented in consciousness, analogous to an exposed film that has not been developed. According to a "law of exclusion" formulated by Poetzl, only the registered but consciously unrepresented parts of a stimulus appear in the dream. The fully "developed" perception, or "print," is excluded from the dream process.

Both Poetzl and Freud were aware of the parallel between Poetzl's findings and Freud's theory of dreams. The "indifferent" perceptions that Freud identified as day-residues present in the manifest content of dreams, Poetzl equated with the unreported parts of the picture which were later dreamed about. Poetzl further assumed that there was a functional equivalence between his concept of the "motoric barrier" and the concept of a repressed wish. He maintained that the dream lifted this "motoric barrier" and permitted the successive development of the parts of the picture originally unseen, much as the dream lifted the barrier of repression and permitted the symbolic and disguised expression of unacceptable wishes. The dream work undertook its own kind of "dark room" procedure, developing the exposed but heretofore undeveloped and unprinted "film" according to its own laws rather than the laws of the waking state.

Recently, Fisher (4, 5, 6) revived general interest in Poetzl's work by providing examples of Poetzl's findings on the basis of which he suggested modifications of Freud's theory of dream formation. Klein and his associates (9, 18) felt the need for a more exact method than Poetzl's procedure, and using an ingenious "masking" technique, have attempted to show that an apparently unseen stimulus affects the perception of a simultaneously perceived stimulus. The work of these two investigators provided the immediate stimulus for the present study. Like Fisher, however, we felt the need to use the opportunities offered by the Poetzl method for tracing the vicissitudes of preconscious percepts from the time of their registration to their emergence in dreams. In a previous paper we reported on a preliminary examination of the Poetzl phenomenon together with an attempt to understand it theoretically (12).

If the basic Poetzl observation could be systematically confirmed and if a measure of delayed recall in dreams could be developed, a new approach to the experimental study of unconscious phenomena would be made possible. This method might successfully handle some of the major difficulties encountered in

[1] We wish to thank our colleagues for critical reading of the manuscript: Gardner Murphy, Riley Gardner, Philip Holzman, Richard Siegal, and Sidney Smith.

the many studies concerned with measuring unconscious phenomena. These investigations —carried out under such headings as subception (11), and perceptual defense and vigilance (3, 10, 15)—by and large, have utilized indirect indices such as variations in threshold time. As Klein and his associates have pointed out, in these studies ". . . non-conscious perception and the defensive reaction to it [are] both inferred from the deviant recognition times. The threshold technique of these studies . . .[provides] little if any information concerning the processes that determine the threshold measure" (9). These studies focus mainly on "defensive reactions to subliminal perception" and not upon the ways in which "subliminal registration affects conscious thought" (9).

Klein *et al.* have taken an important step toward dealing with the problem they pose. Their results suggest another question: what is the *fate* of the preconsciously registered impressions after the moment at which they affect simultaneously evoked conscious impressions? The technique discovered by Poetzl and further developed by Fisher is perhaps best designed to answer this question, for the dream would appear to bring to the surface many of these preconscious perceptions.

The Poetzl procedure has been repeated in several studies (1, 4, 5, 6, 13, 14) but thus far no measure of preconscious perception has been developed. Our primary aim in the present report is to construct a reliable measure of preconscious perception as elicited by the Poetzl technique and then to examine evidence bearing on its validity.

METHOD

Our procedure was essentially the same as the one used by Poetzl and Fisher. They varied conditions unsystematically in what were principally clinical approaches (8). We have attempted to collect our data in a systematic fashion, so that statistical procedures could be applied to the data.

We used the same picture for all *S*s. It provided a baseline from which could be assessed the nature of the "picture pull," as well as individual differences in response to the same stimulus. None of the other experimenters (1, 4, 5, 6, 13, 14) used one picture for a sizeable number of *S*s.

A single description and drawing of a complex picture may not capture all of what the *S* has consciously perceived. Our instructions therefore emphasize that the *S* should describe and draw what he saw as fully as possible. In addition, we asked for a second full description and drawing immediately following the first. At the next session (the following day) we obtained a third sample of the conscious memory of the picture. In order to make certain that nothing would be missed, a tape recording was made of the experimental sessions.

We became convinced of the importance of using a complex picture with many objects to assure that even *S*s who reported a great deal initially would have a sizeable remainder to recover in their dreams. This would make possible a statistically adequate range of recovered items.

It would be foolhardy to expect dreams to occur and be reported on request. We decided to obtain from those *S*s who did not report dreams something similar to a daydream. We asked these *S*s to describe and draw the first mental picture that came to mind. A similar technique was employed by Allers and Teler (1) and more recently by Fisher (5).

The 27 *S*s who took part in this study had participated in other experiments on perception at The Menninger Foundation. Our initial explanations of the nature of the experiment could be brief: "This is a study of perception, or how people see the world. It is similar to some of the studies you've taken part in before." The verbatim experimental instructions follow:

Part I (First Session)

1. On the count of three I'm going to flash a picture on the screen. It will appear just about where the x-mark is. [The experimenter also indicates the size of the area in which the picture will appear on the screen.] Ready: one, two, three. [The picture is flashed on the count of three.]

2. Now, describe what you saw as fully as you can. Make a drawing of what you remember seeing and continue to describe it while you draw it.

3. [The *S* is asked to describe and draw the picture again, as in #2 above. After completing the drawing and description, he is asked,] Is there anything else you remember?

4. Would you remember and write down upon awakening any dreams that you have tonight? Please bring them with you when you come tomorrow. [If the *S* asks, "What if I don't dream?" he is told, "If you keep a paper and pencil handy we've found that this helps a great deal." If he then asks, "But what if I don't dream at all?" the experimenter replies, "Even people who don't usually dream find that this helps."]

Part II (Second Session)

5. Tell me your dream (dreams). Draw and describe what was pictured in your dream, what you saw in your dream.

Alternate instructions for Ss who do not report dreams:

What were the last thoughts you had before falling asleep last night? [This is followed by instruction #6 below. The remainder of the instructions, from #7 on, apply to these images as if they were dreams.]

6. Close your eyes, let a picture come to mind, let me know when you have one. . . . Fine, tell me about the picture that came to mind. Now draw what you saw and describe it as you draw it. [Two pictures are elicited in this way.]

7. [Associations are obtained to the dreams and images by means of the usual questions, "What comes to mind about the dream?" or "What comes to mind?"

—about any part of the dream which the experimenter has reason to think is likely to be important.]

8. Do you see any connection between your dream (image) and the picture you saw yesterday?

9. Now describe what you saw yesterday as fully as you can. Make a drawing of what you remember seeing and continue to describe it while you draw it.

10. I'm going to expose the picture again. [The experimenter exposes the picture on the screen and lets it stay there for the rest of the session.] What connections do you see between your recall of the picture and the picture on the screen? This time you may inspect it at your leisure.

11. What connections do you see between your dream (image) and the picture on the screen? [After this general inquiry, specific questions are asked about parts of the picture.]

12. Do you see any connections between the dream (image) and anything in the room or anything you might have seen on the way down here yesterday?[2]

13. How does this dream compare with most other dreams you've had?

14. What do you think we have been trying to get at in this experiment?

Apparatus

A slide projector with attached shutter was used. The projector was a La Belle, Model 500 with a 500-watt bulb. The shutter was an Ilex, No. 4 Acme with a range from 1 to $\frac{1}{150}$″. Electronic calibration of the shutter was obtained for the exposure time of the experiment, $\frac{1}{60}$″.[3] The screen used was a piece of white poster cardboard 44″ x 28″, and mounted on plyboard. The screen was hung with the long axis parallel to the floor. The room was approximately 9′5″ x 10′8″. The *S* sat against the long wall approximately 9′ from the screen. The pre-exposure field was measured at .005 foot-candles during the daytime and .000 foot-candles at night. At the moment of exposure the illumination was .200 foot-candles. With the room lights on the illumination was at .800 foot-candles. During the pre-exposure period approximately one minute was allowed for adaptation to the new level of illumination. An electronic photometer, Model 500 with a tube D was used for all measures of illumination. The size of the slide on exposure was approximately 21″ x 15″ and appeared on the screen so that the mid-point of the slide was approximately 7½″ above eye level. A black and white reproduction of our colored slide, "Inside Ancient Roman Wall,"[4] can be found in our earlier article (12). A penciled x-mark approximately ½″ x ½″ was drawn on the screen so that it would fall at the exact center point of the picture. *S* was told that the picture would appear "just about where the x-mark is" and the approximate size of the exposed picture on the screen was indicated. *S*'s head was not held stationary. On exposure of the picture the x-mark did not obscure any objects because it fell on a dark green tree in the center of the picture.

A Measure of Recall of Preconscious Perceptions

Basic to our understanding of the Poetzl phenomenon is the assumption that more of the stimulus registers than the *S* can report to us. We shall refer to those elements of the stimulus that register in the *S*'s perceptual system and are not consciously recalled, but later appear in dreams and images, as preconscious. We say recall rather than perception because in its rigorous meaning perception refers to the experience of a stimulus present to the senses; the stimulus in our experiment is only perceptible for $\frac{1}{60}$″. It would be more accurate to refer to our *S*s' reports as the recall of a quite brief perception—a recall in which memory and judgment have already played a role. The dreams and images could also be put in this same frame of reference as another mode of recall, much as we would find in experiments on the hypermnesic character of hypnotic states (17). The recall of the stimulus present in the dreams and images differs from the conscious recall of the stimulus in that the *S* does not have the conscious intent to recall parts of the stimulus and is usually not aware that he is recalling something already exposed to his view. We shall refer to recall present in dreams and images as ***unintentional recall*** in contrast to the ***intentional recall*** obtained immediately after the exposure of the stimulus and on the *S*'s return the following day. (See Table 1.) (By referring to recall in dreams as "unintentional" we do not imply that there are no intentions other than the conscious intent induced by our instructions.)

The ***unintentional recall*** in dreams and images consists of at least two different types of elements: (*a*) Parts of the stimulus already reported in ***intentional recall*** of the stimulus. We shall refer to these as items ***recalled again***. (*b*) Parts of the stimulus previously unreported in ***intentional recall***. We shall refer to these items as instances of ***preconscious recall***.

As a residual category, we should expect to find in the dreams and images items not present at all in the stimulus. We shall refer to these as ***unrelated items***.

Once we view the Poetzl phenomenon as a function of recall we must design the measure of ***preconscious recall*** to take into account the effects of memory, both as a factor producing a decrement in the amount available to recall and as a factor introducing new elements in the form of additions, distortions, assimilations, etc. According to Freud's theory of dreams, we should expect these additive and distorting effects of memory to be more prominent in dreams and images than in ***intentional recall***. As a way of taking into

[2] This instruction is one of the additions introduced by Fisher (4).

[3] We had intended using the same $\frac{1}{100}$″ time exposure as Poetzl and Fisher, but found by electronic calibration of our shutter (Ilex, No. 4 Acme, Ilex Optical Co., Rochester, N. Y.) that the actual time exposure when set at $\frac{1}{100}$″ was $\frac{1}{60}$″. (By courtesy of Charles Snyder.) These types of shutters are inexact, but our time exposure proved consistently to be around $\frac{1}{60}$″. Thus, we succeeded in our aim of providing each *S* with about the same stimulus conditions in terms of length of exposure. This possible difference in time exposure from that of the Poetzl and Fisher experiments (none of the earlier experimenters report on exact calibration) may not be crucial, as is suggested by the work of Fisher (4), and Malamud and Linder (14), both of whom obtained the phenomenon with exposures of half a minute.

[4] National Picture Slide Co., N. Y. C., Slide #9, Italy, 80.

TABLE 1
TIME SEQUENCE OF EXPERIMENTAL PROCEDURES

Experimental Procedure	Time after Exposure
Exposure of picture	0
First intentional recall	immediately following exposure
Second intentional recall	3–5 minutes
Dream (unintentional recall)	10–24 hours
Image (unintentional recall)	24–30 hours
Third intentional recall	24–30 hours

account the transformations in perceptual structure produced by memory, we decided to identify instances of *preconscious recall* in terms of a conceptual similarity between dream and picture elements. Thus, if *S* omitted the trucks from his intentional recall of the picture but dreamed about cars, wagons, or bicycles, we would credit him with the preconscious recall of the vehicles in the picture. We assume that the meaning of the percept—in this case "mode of transportation"—is more constant than any of its visual components, which might have been affected by the particular memories of the individual.[5]

The actual score for *preconscious recall* received by the *S* represents the number of elements in his dream or image having a conceptual similarity to elements of the picture which *S* did not describe or draw in his three intentional recalls. The scores for the other measures (amount of intentional recall, number of items recalled again, and number of unrelated items) were also based upon conceptual categories rather than perceptual identities.

RESULTS

Reliability of the Measure of Preconscious Recall

From the *S*s' descriptions of the picture, we made up a list of 34 major items. The major elements of the picture to which our *S*s most frequently responded were the sky, the statuary on top of the building, the "tower" (facade of the building), the building itself, the trees and shrubbery on either side of the building, the wall on which the building stood, and three vehicles on a roadway at the base of the wall. A few of the reported items were indistinct and infrequently seen and some of them were subdivisions of these principal elements. For example, the building was not only scored as a building, but windows and doors were scored separately when they were drawn and named by the *S*.

Two judges worked independently in scoring the dreams and images according to the presence or absence of the 34 items. The judge was only required to decide whether a listed item was recognizably present in the descriptions and drawings. "Judge 1" represented the pooled judgments of the two experimenters, and "Judge 2"[6] was the statistician who did most of the analyses reported in this article. We were concerned with three scores for each *S*: (*a*) the amount of *intentional recall*; (*b*) the amount of *preconscious recall*; and (*c*) the amount *recalled again*. The reliability coefficients (Pearson *r*'s) for these three variables were respectively: .92, .84, and .95.[7]

Establishing the Relationship of the Measure to Preconscious Perception

Like Poetzl, we found that parts of the picture which were not reported as seen later appeared in the dreams. Out of an average of about ten items appearing in dreams, three had not been reported previously (see Table 2).

[5] Another approach to the problem of identifying instances of *preconscious recall* is to conceive of a preconsciously registered percept as a dissolving photographic image whose fragments become incorporated into the imagery of dreams. From this point of view, there need not be any conceptual similarity between *preconscious recall* and the unreported aspects of the stimulus but only a fragmentary perceptual identity. Fisher relied primarily on this method of identifying *preconscious recall*. There probably is a correlation between these two aspects of *preconscious recall*. In our previous illustration concerning the omitted trucks, the conceptual similarity between the dream and the stimulus is clearly that of a mode of transportation or vehicle; but there is also a perceptual identity present insofar as an *S* usually draws his wagons with wheels, fenders, etc. Our decision in favor of a measure based on conceptual similarity was a first step dictated by the need to achieve adequate reliability, which seemed more readily obtainable by judging conceptual similarities than by judging minute, subtle, and fragmentary perceptual identities.

[6] Lolafaye Coyne.

[7] The score given each *S* by the two judges could be based on different items. We therefore determined the amount of agreement between Judge 1 and Judge 2 for all *items*. This is the same as asking how well the judges agreed on whether an *S* was drawing and describing, for example, a building, a car, or a tree. The percentage agreement between the two judges for all items, pooling all *S*s, was 68. Almost all the disagreements had to do with one judge scoring an item and another not scoring it. There were two main reasons for these disagreements. Occasionally, one judge noticed an ambiguous item and decided it did not represent the object, while the other judge decided that it did resemble a recognizable object. A more prevalent type of disagreement was essentially one of oversight by one or the other of the judges of an unambiguously represented item. The descriptions were rated more unreliably than the drawings. When oversights were called to the attention of the judges they usually immediately confirmed having overlooked the object.

We also found that another source of *unintentional recall*, images, was effective in eliciting new picture items. Indeed, the images proved to be significantly more productive of *preconscious recall* than the dreams (see Table 2), a finding that raises some interesting questions beyond the scope of this paper. We have labeled these new items in the dreams (and images) *preconscious recall*. How can we show that this is a proper description?

If we could put "tracers" on the picture items analogous to radioactive isotopes, we could follow them through the mental apparatus and our task would be relatively simple. Until such psychologically radioactive isotopes are discovered, however, we must learn about the validity of the measure by bringing together evidence that it behaves in accordance with our understanding of preconscious perception. We explored hypotheses based on the assumption that preconscious perception occurred and was utilized in dreams and images.

Poetzl's "law of exclusion." Poetzl stated: "... out of the total effect of a visual situation the once-performed part [i.e., the part which appeared in consciousness] disappears in the after-effect, but those preconsciously grasped parts, which were not yet developed, continue to be effective. If the grasp of consciousness is at its maximum, the latent after-effects in preconsciousness are at a minimum and vice-versa" (16). This "law" might lead one to expect that the more items the *S* reported (*intentional recall*), the smaller the number of items left to be "developed" in the dreams and the images (*preconscious recall*); similarly, the greater the *S*'s "conscious grasp" of the picture (*intentional recall*), the less the number of items *recalled again* in the dreams and images. Our findings indicated otherwise. The correlation between *intentional recall* and *preconscious recall* was not significant (−.06), while the correlation between *intentional recall* and *recalled again* was significant, but in the opposite direction from that predicted (.69, $p < .001$). Poetzl therefore was wrong if the two predictions followed from his "law of exclusion"—the amount one sees initially of a picture is unrelated to the amount one recovers of it in dreams and images; and the more one sees initially, the *more* often the same things will be seen again in the dreams and images.[8]

TABLE 2

MEAN NUMBER OF ITEMS RECALLED IN EACH CATEGORY FOR DREAMERS AND NONDREAMERS

Category of Recall	Dreamers (N = 16)	Nondreamers (N = 11)	t
Preconscious recall	2.9	4.7	2.25*
Intentional recall	9.3	9.5	0.17
Items recalled again	2.4	3.6	1.09
Unrelated items in unintentional recall	4.7	5.5	0.44

* $p \leq .05$.

In view of these negative results, one may wonder whether we are measuring the same phenomenon that Poetzl observed. There is, however, another way of testing Poetzl's "law of exclusion" made possible by our obtaining three intentional recalls of the picture, rather than one, as in previous studies. Quite frequently we noted that items not in the first recall emerged in the second or third recalls. We were led to ask whether more new items would appear in the dreams (and images) than in successive intentional recalls. Possibly Poetzl's "law" was intended mainly to emphasize that the dream is a better vehicle than *intentional recall* for *preconscious recall*, rather than to imply that the dream is the exclusive vehicle for preconscious perception. Our definition of *preconscious recall*—the appear-

[8] The relationship between *intentional recall* and *preconscious recall* could be obscured by an uncontrolled factor—we have no way of knowing how much of the picture actually registered in our *S*s' visual apparatus. There may be wide individual variations in how much a person can "receive," adding together conscious and preconscious perception. Visual acuity might be one such variable which would place a physiological limit upon how much could be perceived at any one time. An analysis which pitted the *S* against himself would control for inter-individual variations in visual acuity and other limiting physiological conditions.

The relationship between *intentional recall* and the number of items *recalled again* in dreams and images is also limited by an uncontrolled factor: on the basis of chance expectations the more things a subject saw in the picture consciously, the greater number of things he could recall again in his dreams or images. If a person reported seeing ten items, he could see again in his dreams a maximum of ten items; but a person who saw only two items could only recall a maximum of two in his dreams. In this sense, a positive correlation was "built into" our measure. If the *S* were compared with himself under the conditions of *intentional* and *unintentional recall*, the amount consciously reported would be eliminated as a source of individual variation.

ance in dreams and image of picture items not previously reported in *intentional recall*—might therefore have been too limiting. Perhaps the new items in successive *intentional recalls* could be viewed as the feeble workings of *preconscious recall* under the adverse conditions of the waking state. Similarly, perhaps the dreams and images are relatively inefficient means as compared to *intentional recall* for recalling conscious perceptions. We, therefore, restated our two predictions:

1. Each *S* should recall significantly more new items (*preconscious recall*) in his dreams or images than in his second and third efforts at *intentional recall*.

2. Each *S* should recall significantly fewer old items (*items recalled again*) in his dreams or images than in his second and third efforts at *intentional recall*.

Both predictions were borne out by our results. (See Tables 3 and 4.) For our analyses, a difference score was obtained for each individual between his performances in *intentional* and *unintentional recall*, using the first intentional recall as a baseline. In analyzing *preconscious recall*, for example, the difference was obtained between the number of new items in *unintentional recall* (dreams or images) and the number of new items present in *intentional recall* (second or third recall, or their average, depending on the particular comparison involved). These difference scores were averaged for all *S*s and the significance of the average difference tested. *Intentional recall* was a better vehicle for retaining conscious perceptions, but could not measure up to the capacity of *unintentional recall* to activate preconscious perceptions. These relationships applied to dreams and images taken separately or combined. The findings support our restatement of Poetzl's assertion that dreams are prone to "exclude" conscious perceptions in favor of preconscious perceptions. And, our results indicate that this "law" can be generalized to apply to a type of "forced fantasy," the images.

Freud's hypothesis about "indifferent" perceptions. We wondered what characteristics of preconscious perceptions made it possible for them to play a role in dreams and images. In his theory of dreams, Freud identified preconscious perceptions as those which are "... indifferent and thus had no attention paid to them; ... or have been rejected and have thus had attention promptly withdrawn from them" (7). In either case, these preconscious impressions "have not yet been requisitioned by waking thought-activity; and for reasons of censorship ... [the dream process] transfers psychical intensity from what is important but objectionable to what is indifferent" (7). In this last statement Freud suggested a reason for the operation of Poetzl's "law of exclusion." Fully developed conscious perceptions cannot so easily be used as a "cover" for unacceptable thoughts.

We would predict then that preconscious perceptions, as they emerge in dreams, would

TABLE 3

COMPARISON OF PICTURE ITEMS PRECONSCIOUSLY RECALLED IN UNINTENTIONAL AND IN INTENTIONAL RECALL

Comparison	*N*	Mean Difference	σ_D	*t*
Unintentional recall vs. second intentional recall	27	2.3[a]	2.75	4.35****
Unintentional recall vs. third intentional recall	26	2.5	2.40	5.50****
Unintentional recall vs. second and third intentional recall combined	26	2.4	2.52	4.93****
Second intentional recall vs. third intentional recall	26	0.3	1.55	1.03

**** Significant at $p < .0001$.

[a] A positive difference is in favor of the first variable mentioned in the comparison.

TABLE 4

COMPARISON OF PICTURE ITEMS RECALLED AGAIN IN UNINTENTIONAL AND INTENTIONAL RECALL

Comparison	*N*	Mean Difference	σ_D	*t*
Unintentional recall vs. second intentional recall	27	−2.9[a]	2.49	−6.02****
Unintentional recall vs. third intentional recall	26	−3.0	3.03	−5.16****
Unintentional recall vs. second and third intentional recall combined	26	−3.0	2.67	−5.81****
Second intentional recall vs. third intentional recall	26	−0.0	1.90	−0.11

**** Significant at $p < .0001$.

[a] A negative difference is in favor of the second variable mentioned in the comparison.

be associated with a strong affective charge displaced from ideas originally intensely cathected to ideas having only a weak charge. Moreover, because these intensely cathected ideas are the derivatives of unacceptable unconscious wishes, we should expect the affective charge to be unpleasant in nature. With these considerations in mind, we predicted that *preconscious recall* would be associated with unpleasant affect.

In order to check this prediction, we devised a measure of unpleasantness using a five-point scale in which two judges rated the maximum point of unpleasantness attained (*peak unpleasantness*) within the dreams, images, and the three reports of *intentional recall*. One judge was not aware of the expected relationship between *peak unpleasantness* and *preconscious recall*. We did not get the *S*s' own rating of the dream's unpleasantness. Illustrating a low *peak unpleasantness* rating would be a dream in which there was no more unpleasant part than a man reading a book; illustrating a high rating would be a dream in which a woman was sitting in her car while terribly frightened by a hail storm.

Confirming our prediction that *preconscious recall* would be associated with unpleasant affect, we found a positive correlation between *peak unpleasantness* and the amount preconsciously recalled. Both for our group of *S*s as a whole (.57, $p < .01$) and for the dreamers (.75, $p < .001$) and nondreamers (.60, $p < .05$) the relationship was significant.

It could be argued, however, that the unpleasantness in dreams and images may not be a function of the preconsciously recalled items at all, but a function of other aspects of the dream. With this objection in mind, we correlated *peak unpleasantness* with the number of items that were *recalled again* and the number of *unrelated items* in the dreams and images. The two correlations were, respectively, .01 and .29. These nonsignificant correlations support the assumption that it was the preconsciously recalled items and not the other contents of the dreams and images that were associated with *peak unpleasantness*.

An hypothesis concerning a formal attribute of the picture. Our measure of *preconscious recall* was based on the picture's content. The *formal* attributes of the picture may also be capable of *preconscious recall*. One formal property is the angularity of line produced by the shapes of the picture items.

Each drawing made by an *S* was judged according to its *angularity* (as opposed to roundedness) on a five-point rating scale. For the drawings of the picture a reliability coefficient of .88 was obtained for two judges; the reliability coefficient for the dream and image drawings was .81.

In constructing our measure of *angularity*, we assumed that the average angularity for the group of *S*s would be a good estimate of the angularity of the picture. Statistical analysis indicated that there were no significant differences among the means for *intentional* and *unintentional recall*, a finding in line with the assumption that the mean remained a relatively constant estimate of the angularity of the picture. We can consider individual deviations from the central tendency as errors in reproducing the angularity of the picture. A "z" score was obtained for each *S* representing the difference between the angularity of his *unintentional recall* drawings and the mean angularity of the total group for *unintentional recall*. We would expect that the smaller the "z" score (i.e., the closeness with which the *S* approximates the average angularity of the drawings), the greater the amount of *preconscious recall* of picture contents. The correlation between the "z" scores and *preconscious recall* of picture contents was $-.49$ ($p < .01$). This relationship was stronger for the dreamers ($-.57$, $p < .05$) than the nondreamers ($-.22$).

Preconscious recall of the picture content was shown to be significantly correlated with *peak unpleasantness*; angularity should similarly be correlated. We found a significant correlation in the expected direction ($-.40$, $p < .05$).

As with *peak unpleasantness*, however, it could be argued that *angularity* was equally related to other aspects of recall, such as items *recalled again* and *unrelated items*. Our expectation was that there would be no systematic relationship between *angularity* and these two variables. The correlation between angularity of dream and image drawings and the number of items *recalled again* was $-.09$. The correlation between *angularity* and the number of *unrelated items* present in the dreams and images was, however, significant, $-.41$ ($p < .05$), indicating that the nonpicture items

appearing in *unintentional recall* are in some way related to the accuracy with which the angularity of the picture is represented in the dream and image drawings.

In view of these findings, the angularity of the picture can perhaps be treated as an aspect of *preconscious recall*. The degree to which this angularity is rendered accurately is related to *peak unpleasantness* and items *recalled again* in the same way as the *preconscious recall* of picture contents. Our measure of *preconscious recall* relied on the conceptual similarities found in the picture content. *Angularity*, on the other hand, would extend our measure of *preconscious recall* to perceptual identities between the dream and picture.

Evidence Against the Chance Occurrence of the Picture Items in Dreams and Images

One disadvantage of our measure of *preconscious recall* derives from the range of items that can be included in a conceptual category. How certain can we be that the wagons seen in the dream, for example, are not derived from a source other than the trucks in our picture? And when we note that most of the objects in our picture are commonplace, the strength of this criticism is increased. If we could, however, select from our pictures the items that occur rarely in everyday experience and demonstrate that these rare items are preconsciously recalled more frequently than would be expected by a standard of chance occurrence, we would have presumptive evidence that the source of *preconscious recall* was our picture and not an extraneous stimulus. Comparison data from the literature on extrasensory perception may help establish a standard by which to judge frequency of occurrence of percepts. Catalogues have been constructed to show how frequently an *S* will guess an unseen object, selected at random and without his knowledge. A catalogue has been made for British *S*s (2), and now an unpublished version for American *S*s by Taves (19) (available by courtesy of Gardner Murphy) of 8,164 items produced by 272 *S*s.

The Taves catalogue can help us answer this question: are the items in our picture that have a low frequency of occurrence in the Taves catalogue and are presumably uncommon in everyday experience recalled preconsciously more frequently than their rare occurrence in the Taves catalogue would lead us to expect? If we find, for example, that an item in our picture which occurs only two per cent of the time in the Taves catalogue, but occurs 20 per cent of the time in our *preconscious recall*, it would seem likely that the picture and not some extraneous source contributed that item. When we selected those items in our picture which made up two per cent or less of the total number of responses in the Taves catalogue, we found that when added together those same items occurred much more frequently in preconscious recall. The chi square was at the .008 level of significance. (If we selected items which occurred .05 per cent or less of the time in the Taves catalogue, the results were even more significant—the chi square was below the .001 level of significance.)

The comparison with the Taves catalogue supported our assumption that nonexperimental sources alone cannot account for *preconscious recall*. For future work, however, it would be helpful to compile item frequencies for dreams and images, so that the standard of chance occurrence might be more closely related to our data than the ESP catalogues.

DISCUSSION

Our findings form a pattern consistent with the assumption that preconscious perception exists and can be elicited and measured. We have developed a measure, *preconscious recall* in dreams and images, which behaves in accordance with an understanding of preconscious perception. Our *S*s' dreams and images contain significantly more originally unreported items than do successive efforts at *intentional recall*, while retaining fewer of the originally seen parts of the stimulus. An estimate of unpleasant affect is associated with the new items appearing in the dreams and images. These two findings support Poetzl's "law of exclusion" and Freud's conception of preconscious perception as the raw material "preferred" by the dream. The finding concerning the *angularity* of the drawings indicates that the measure of *preconscious recall* can perhaps be extended to such formal properties of the stimulus as shape and position. A method was successfully applied for assessing the "chance" occurrence of picture items in dreams and images. The rare items in the picture appear significantly more fre-

quently in dreams and images than a standard of expected chance occurrence (i.e., Taves catalogue). Taken together these four findings strengthen the conviction that Poetzl's technique for relating dreams to specific and identifiable daytime stimuli is an effective method for the experimental study of dreams.

Much more remains to be understood about the phenomenon. For example, why are the *unrelated items* unassociated with unpleasant affect? Many of these unrelated items could be day residues stemming from nonpicture sources, and as such should have unpleasant affects associated with them.

It would be desirable if a method could be devised for identifying at what point in the cognitive process the percept acquires anxiety arousing connotations and becomes available as a day residue. Customarily, it has been assumed that anxiety arousing connotations are present at the moment of perception and indeed produce the perceptual omissions or distortions. The process may be the reverse—an initial ambiguity or vagueness in the percept may make it possible for anxiety arousing meaning to adhere to it at some later point.

Lastly, the focus of our measure, preconscious perception, is upon items omitted from conscious report. But many items that were reported as seen appeared again in the dreams and images. Do items that are reported as seen appear in the dream via a different process from omitted items? A thorough study of the *S*'s initial reports may reveal many different types of items which are as likely to appear in dreams and images as initially omitted items. We have preliminary evidence that doubt surrounding items is one of the most significant indications that they will appear in dreams and images.

Among other factors to be taken into account in future experimentation, it would be naive to neglect the fact that what is reproduced in *preconscious recall* is stimulated by much more than the experimenter's picture. Frequently the contents of the experimental room and the building become part of the "experimental stimulation." There was at least one striking example of this. The masonry which was occasionally correctly located in the wall below the building in the picture could as easily refer to the walls of the office in which the experiment took place, which were of unplastered concrete blocks.

It could be argued that the *amount* of *preconscious recall* is a function of the number of items the person reports in his dreams and images. If the *S* talks long enough he may mention many of the common items included in our picture. The simplest way to test for this is to count the number of items unrelated to the picture appearing in the dreams and images and to correlate this sum with the sum for *preconscious recall*. The correlation for the entire group was not significant (.29).

Several methods will be profitable for future studies. (*a*) *Prediction study:* By predicting from the *intentional recall* of the picture what items will appear in a dream (or image), we have another way of exploring the phenomenon. By obtaining more than one complete description and drawing of the picture in quick succession, predictions are facilitated. We observed that changes from the first to second set of drawings and descriptions were often predictive of what would appear in the dreams and images. Predictions have been tried on an unsystematic basis with encouraging success (12). (*b*) *"Own Control" method:* Images can be elicited both before and after a stimulus picture is exposed. A comparison of the images produced before seeing the picture with those produced afterward can point to the influence of the picture as compared with the readiness for imagining nonpicture items. This procedure has also been tried unsystematically (6). (*c*) *Experimental pictures containing unusual objects:* As the Taves frequencies indicated, most of the objects in our picture are common. It is important to use a picture in which many or all of the objects are rare. The frequency with which the objects are commonly associated with each other should also be controlled. Instances of *preconscious recall* might really be commonplace associations—car and street, shrubbery and house, building and windows, etc. The new items in the dreams and images could be *associations* to the originally seen items, dream continuations of thoughts begun in a conscious state. (*d*) *Intra-individual consistency in response to a number of pictures:* By using a number of different pictures for the same person we can separate effects due to individual differences from the effects produced by picture content. The other three

methods can be employed with a number of pictures as easily as with one.

SUMMARY

In a pioneer investigation, Poetzl discovered that *S*s dreamed about the originally unreported parts of a picture exposed tachistoscopically. If this discovery can be confirmed and a measure developed of preconscious perceptions as they appear in dreams, a new approach becomes available for the experimental study of unconscious phenomena.

A kodachrome slide was exposed at $\frac{1}{50}$″ to 27 *S*s. They were asked to describe and draw the picture three times and then to record any dreams they had that night. From *S*s who did not report dreams an "image" was obtained by asking them to close their eyes and describe and draw the first picture that came to mind.

High interjudge reliability (.84) was achieved for a measure of preconscious perception (*preconscious recall*) based on the number of new picture items appearing in the dreams and images. The meaning of the measure was explored by testing two hypotheses:

1. Poetzl's law of exclusion: Dream imagery "excludes" conscious perception in favor of preconscious perception.

2. Freud's hypothesis about "indifferent" perceptions: Freud asserted that the neutral or "indifferent" character of preconscious perception permitted it to serve as a "cover" for unconscious ideas which would not otherwise escape dream censorship. If preconscious perceptions were the ones that most often linked up with derivatives of threatening ideas, unpleasant affect would more likely be associated with the recall of preconscious perceptions than with other aspects of the dream.

Our data very significantly supported the Poetzl hypothesis for picture content and also for a formal attribute of the picture, its angularity. Consistent with our inference from Freud's hypothesis, we found a significant correlation of .57 between *preconscious recall* and *peak unpleasantness*.

These findings provide evidence for the Poetzl phenomenon consistent with the assumption that preconscious perception can be elicited and measured both in images and in dreams.

REFERENCES

1. ALLERS, R., & TELER, I. "Ueber die Verwertung unbemerkter Eindruecke bei Associatonen." *Ztschr. f. Neurol. und Psychiat.*, 1924, **89**, 492–513.
2. CARINGTON, W. Experiments on the paranormal cognition of drawings, III. *Proc. Amer. Soc. for psych. Res.*, 1944, **24**, 1–107.
3. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564.
4. FISHER, C. Dreams and perception: The role of preconscious and primary modes of perception in dream formation. *J. Amer. psychoanal. Assn.*, 1954, **2**, 389–445.
5. FISHER, C. Dreams, images, and perception: A study of unconscious-preconscious relationships. *J. Amer. psychoanal. Assn.*, 1956, **4**, 5–48.
6. FISHER, C. A study of the preliminary stages of the construction of dreams and images. *J. Amer. psychoanal. Assn.*, 1957, **5**, 5–60.
7. FREUD, S. *The interpretation of dreams*. (Trans. by James Strachey) New York: Basic Books, 1955.
8. KLEIN, G. Discussion remarks on Charles Fisher's "A study of the preliminary stages of the construction of dreams and images." Unpublished manuscript read at the New York Psychoanalytic Society, 1956.
9. KLEIN, G., SPENCE, D. P., HOLT, R. R., & GOURE-VITCH, S. Preconscious influences upon conscious cognitive behavior. Paper read at Annual Meeting, American Psychological Association, San Francisco, 1955.
10. LACY, O. W., LEWINGER, N., & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.
11. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **48**, 113–122.
12. LUBORSKY, L., & SHEVRIN, H. Dreams and day-residues: A study of the Poetzl observation. *Bull. Menninger Clin.*, 1956, **20**, 135–148.
13. MALAMUD, W. Dream analysis: Its applications in therapy and research in mental diseases. *Arch. Neurol. & Psychiat.*, 1934, **31**, 356–372.
14. MALAMUD, W., & LINDER, F. E. Dreams and their relationship to recent impressions. *Arch. Neurol. & Psychiat.*, 1931, **25**, 1081–1099.
15. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
16. POETZL, O. Experimentell erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Ztschr. f. d. ges. Neurol. und Psychiat.*, 1917, **37**, 278–349.
17. ROSENTHAL, B. G. Hypnotic recall of material learned under anxiety and non-anxiety producing conditions. *J. exp. Psychol.*, 1944, **34**, 369–389.
18. SMITH, G. J. W., & HENRIKSSON, M. The effect on an established percept of a perceptual process beyond awareness. *Acta Psychol.*, 1955, **11**, 346–355.
19. TAVES, E. The construction of an American catalogue. *J. Amer. Soc. psych. Res.*, 1945, **39**, 151–156.

Received June 3, 1957.

AFFECTIVE RESPONSE TO PSYCHOANALYTIC DEATH SYMBOLS

W. W. MEISSNER

Saint Louis University

IN RECENT years, an attempt has been made to account for the anxiety reaction typical of neurosis in terms of a "fear of death" (5, 14). Psychiatrists and psychoanalysts have begun to recognize that at least some kinds of neurotic behavior can be described and treated in relation to this dread of death. In general, this fear seems to take the form in which the person dreads the dissolution of his own personality. The present study is an attempt to investigate the possible relation between the death concept and emotional response.

The measure of affectivity employed in this study was the GSR. The general evidence on the GSR as a measure of emotional response has been summarized by Woodworth and Schlosberg (19). Particularly relevant for this investigation is the evidence which suggests that the GSR is applicable as a measure of unconscious emotionality. The work of McCleary and Lazarus (8, 9) on autonomic discrimination without awareness, McGinnies' findings on perceptual defense (10, 11), and Jones's experiments on social behavior in relation to GSR responses (6) seem to indicate that the GSR is an appropriate index of internal, inhibited, unconscious emotionality. The hypothesis of this study is that the presentation of stimulus words, which have been found by psychoanalytic methods to be symbolic of the death concept, elicits unconscious emotional responses, which are indicated by a greater GSR amplitude to such words than to neutral terms.

METHOD

Subjects

The *S*s were 40 Roman Catholic seminarians with considerable training in humanities and religion. Ages ranged from 23 to 45 years, with a median of 28 years. They were all psychologically naive and had no information about the purpose or specific content of the experiment.

Apparatus and Materials

The GSR was recorded on a 110-volt, 60-cycle A.C. Garceau recording galvanometer, made by the Electro-Medical Laboratory. Circular palmar electrodes of amalgamated zinc, applied with Brunswick electrode paste, were used.

Twenty death symbol words were selected from the literature and randomly mixed with 30 randomly selected words from the Kent-Rosanoff Word List (13) to compose the experimental word association list. Ten additional words were drawn from the Kent-Rosanoff list to provide a warm-up period for the *S* to adapt to the experimental procedure. The experimental design is based on the comparison of responses to the experimental grouping of 20 death words with the responses to the 30 control words.

The same 20 death symbol words were also used to compose a written test,[1] in which the words were simply presented in the following order:

1. bird	11. black
2. journey	12. water
3. candle burning out	13. the silent one
4. to burn	14. to depart
5. across the bridge	15. thunder
6. sleeping person	16. stranger
7. train trip	17. terminal
8. old man	18. the end
9. statue	19. four
10. to drive away	20. the thirteenth

*S*s were given the written test with the following instructions, which were printed on the test blank: "This is a test of interassociation. Try to group as many of the words as possible under a common association or concept or common idea to which they refer. Then indicate the common idea, together with the numbers of the words from the list which you associate with it, in the spaces provided. Do this for each common association. The more common associations you can find and the more words you can relate to each association, the better it will be." This test was used to verify the content of the responses to these death symbol words when they were presented in the experiment.

Procedure

The experimental procedure (16) was the same for each of the 40 *S*s. The *S* was brought to the experimental room and asked to lie down on a bed provided for the purpose. He was told only to relax and rest. The rest period lasted for 15 minutes; during this period, the galvanometer was permitted to warm up. The electrodes were then attached. A small amount of nonpolarizing electrode paste was applied to the *S*'s palm, and a small amount of paste was evenly spread over the surface of the electrode. The electrode was held in place by an elastic band.

The *S*s were instructed to (*a*) respond to the stimulus word with the first word that they happened to associate with it, (*b*) reflect quietly on the content of their response, (*c*) remain as relaxed and inert physically as possible, and (*d*) reflect only on the content of their response, excluding all other extraneous ideational

[1] This written test was empirically constructed on another group of 20 seminarians comparable to the experimental subjects.

material. Then the ten warm-up words were presented, followed immediately by the 50 experimental words. The words were pronounced in an even and neutral tone of voice. A reading was taken on the galvanometer at the point of the presentation of the stimulus word and at the point of maximal deflection for each response. The difference between these two readings was taken as the change in resistance for that particular response. At each stimulus, *E* waited until the level of resistance had returned to a relatively steady condition before presenting the next stimulus. The interval between stimuli varied from word to word and from *S* to *S*, but the average interval was approximately 20 seconds (2). The entire procedure lasted about 50 minutes for each *S*.

In the computation of the results, some of the changes from the first to the second reading which were positive instead of negative, that is, which indicated a rise in resistance instead of a drop, were eliminated by equalling them to zero. It seemed that these positive responses were an artifact of the constant positive drift observed in all *S*s. Evidently, these responses were composed of a small negative decrement which was immediately overcome by the positive drift and thus could not be recorded by *E*. This happened in only a few instances.

As soon as the list of stimulus words had been presented and the GSR responses recorded, the electrodes were removed. The *S* was then asked to write down all the stimulus words (those pronounced by *E*) that he could recall. No time limit was imposed. It was thought that a significantly higher recall of death stimulus words would indicate greater affective involvement. The differences, however, were not significant. This would seem to support Merrill's conclusions in regard to the inadequacy of recall for indicating affective involvement in word association procedures (12).

Immediately after this, the *S*s were given the written test described above. *S*s were told that they could take as much time as they pleased in filling out the test. The only other instructions were those printed on the test blank. The written test was constructed to show that the responses to the death symbol words, which were measured by the GSR, were made on the basis of death association and not on some other vicarious association. The *S*s responded by writing in the provided space a word which any or all of the death symbol words may have suggested to them. The numbers of the death symbol words which suggested this common referent were recorded together with the referent. A typical response would be: "Death—2, 3, 6, 8, 11, 13, 14, 18." There was no limit to the number of referents an *S* could give. Each such referent was scored as a response, and was given a value score equal to the total number of words associated with it. Thus, the typical response cited above would have a score of 8. The highest possible score for any one response was 20, since there were only 20 words in the list. The total group score for each such referent was obtained by adding the individual scores from each *S*. Only explicit reference to the word "death" was taken as evidence of association to the death concept.

Conditions

Conditions for the experiment were structured with Venables' findings on the effect of temperature and humidity on the GSR in mind (17). The experiment was conducted in a soundproofed, airconditioned room of constant temperature and humidity. The temperature was kept at approximately 72 degrees Fahrenheit, the humidity at 70 to 75 per cent. There were no distracting noises and no interruptions to disturb the *S*s.

RESULTS

The written test of interassociation was intended to yield an index of the tendency of these *S*s to associate the death symbol words to the death concept. In this way it was hoped that some indication could be obtained of the extent to which the GSR responses to the death symbol words were actually determined by association to the death concept. The total score of all the responses given by the 40 *S*s was 2430. Out of this total, the death referent had a total score of 267, which was higher than the score of any other referent. No direct statistical evaluation was made of the above chance expectation of the death referent score. However, a random selection of an equal number of words from the Kent-Rosanoff list provides an association with the death concept in only three out of 20,000 responses (13). A statistical comparison of these proportions is not possible due to the differences in the method of obtaining responses, but the comparison is nonetheless suggestive, if not indicative.

It was also felt that the frequency of the death association as indicated by the death score should be compared to the frequency of association to the other referents given by the *S*s. To obtain this measure, the proportion of death responses was compared by a *t* test with the proportion of that referent which had the second highest score. The referent "travel" was second highest with a score of 200, and comparison with the death score yielded a *t* of 3.265. This difference is significant at well below the .01 level and indicates that other factors than chance were operative. Thus, the death score here suggests that the *S*s to a significant degree associated the death symbol words with the death concept and that these words elicit associations to the death concept with greater probability than to any other referent.

The experimental data obtained from the GSR measurements were treated in a simple analysis of variance for matched groups or repeated observations (3, 7). Because of the high frequency of associations to the travel concept, it was felt that the analysis should be

conducted on three groupings. By an analysis of the data from the written test, the 20 death symbol words were divided into those which seemed related to the travel concept and those which did not. These two groups, together with the neutral words from the stimulus list, constituted the three categories: Death-non-travel, Death-travel, and Non-Death. This design was intended to determine the significance of the differences between the amplitude of the GSR responses to the death symbol words, on the one hand, and the amplitude of the responses to non-death or neutral words, on the other, and whether the words related to the travel concept actually constituted a separate response category from the death association in the GSR measurement. A significant difference between the Death-non-travel and the Death-travel groups would suggest that the responses to each group of words was governed by a different set of associations. The results of the analysis of variance are presented in Table 1.

The significance of the differences revealed by the analysis of variance made the application of the *t* test necessary to locate the differences. The *t* was applied with the error term or residual of the analysis as the variance of the means. Results are presented in Table 2. The difference between Death-non-travel and Non-Death words, and the difference between Death-travel and Non-Death words are significant beyond the .01 level. The difference between the Death-non-travel and the Death-travel words is nonsignificant. Within the limits of this design, one may conclude, then, that the greater amplitude of the GSR response to death words than to neutral words was probably due to association to the death concept, in so far as the *S*s tend to associate these words to "death" more than to any other concept.

TABLE 1

ANALYSIS OF VARIANCE FOR GSR RESPONSES[a] TO DEATH-NON-TRAVEL, DEATH-TRAVEL, AND NON-DEATH STIMULUS CATEGORIES

Source of variation	*df*	Mean square	*F*
Between response categories	2	0.93937	24.323*
Within response categories	117	0.47797	
Between *S*s	39	1.35668	35.129*
Residual	78	0.03862	
Total	119		

[a] Calculations based on log conductance units.
* Values are significant at less than the .01 level of confidence.

TABLE 2

t TESTS APPLIED TO THE ANALYSIS OF VARIANCE FOR GSR RESPONSES[a]

Categories	*t*	*P*
Death-non-travel vs. Non-Death	6.441	<.01
Death-travel vs. Non-Death	5.538	<.01
Death-non-travel vs. Death-travel	0.903	n.s.

[a] Calculations based on log conductance units.

DISCUSSION

The results of the analysis of variance and the *t* tests applied to it indicate that the differences between the amplitude of the GSR in response to death symbol words and the amplitude in response to the neutral control words are significant. This result supports the main hypothesis of this study that death symbols elicit a differential emotional response. The results shown in Table 2 also indicate that responses to the Death-travel words do not differ significantly from those to the Death-non-travel words. There is therefore no evidence that the travel words evoked a distinguishable category of GSR response. The responses to both were apparently based on a common associative content involving the death concept. The reason for the prominence of the travel responses on the written interassociation test seems to be simply the direct implication of travel in a number of the terms.

The hypothesis behind the written interassociation test was that the *S*'s reflection on his own responses to the death stimulus words would tend to tap the complex of associative material which related to the death concept and involved an affective reaction. The presentation of the same stimulus words together in a group on the written test should elicit the same associations as did these same words in the experimental presentation. This multiple probing of the same association area on the written test should yield, then, explicit reference to the death concept if these symbolic words had any tendency to be referred to the same psychic content. The results of the written test indicate that such a tendency was manifested for these *S*s over these symbolic words. This evidence does not directly imply that the responses measured by the

GSR were actually responses involving reference to the death concept. But the tendency of these *S*s to associate these words to the death concept increases the probability that this was in fact the case.

During the course of the experiment, the infrequency of the overt response of "death" to the death symbol stimulus words was noteworthy. When death symbol words were presented, the *S*s responded with the word "death" only three or four times out of 800 opportunities. This low frequency, compared to the much higher one elicited by the written test, seems to suggest that the reaction to the death words which occasioned the greater GSR deflections may well have been on an unconscious level. This possibility fits in well with previous findings on the GSR as a measure of unconscious affectivity. The relation is at any rate highly suggestive.

The results and conclusions presented here presume that the *S*s were able to observe all the experimental conditions. No check can be made with any accuracy on the extent of interference from extraneous ideational material during the course of the experiment. *S*s were directed to exclude such material, but this is no guarantee. A claim could be made for the unconscious character of the connection between symbol and associative pattern, so that one could hypothesize that the *S* was reacting on the basis of death association. However, this hypothesis was not open to validation in this experimental design.

A factor that may have biased the results of this study is the common religious attitudes and training of the *S*s. One might expect a greater involvement in the death concept in persons for whom it has greater meaning. The extent of this influence can only be estimated from parallel studies with an equivalent design on different *S*s. Another uncontrolled factor was the tone of voice in which the stimulus words were presented. An attempt was made to pronounce them in a neutral tone, but the possibility remains that the *S*s picked up some subtle clue which may have influenced their responses.

It should be pointed out that the positive drift of the resistance level, which is commonly observed in *S*s in a relaxed condition (1), may have tended to decrease differentially the amplitude of the responses. This would have a slightly greater effect on the responses of greater amplitude, tending to reduce the greater resistance drops toward the average of the lesser resistance drops. Over the whole spread of measures, this would reduce the mean difference between the death and non-death responses.

McGinnies' demonstration of higher GSR responses to emotionally toned words (10, 11) was criticized by Howes and Solomon (4) on the ground that the emotionally toned words were less frequent in common usage and that this may have accounted for the GSR differences. This same criticism might be applied to the present design. However, checking the frequencies of the death words used in this study against the frequencies of the non-death words on Thorndike's frequency list (15) indicates that there is no significant difference in the relative frequencies.

Methodologically, the device used here to identify associative content is difficult, demanding, and in the final analysis yields nothing but indirect evidence for a plausible argument. It suggests only a tendency in the *S*s and does not identify the precise nature of their responses. Improved instrumentation may well be worth the effort of devising it. Additional light may be thrown on the connection between anxiety and the death concept by the introduction of some clinical criterion or index of anxiety. In this connection, GSR results from normal and neurotic groups might be compared. The effects of induced anxiety by the use of lysergic acid, mescaline, and other anxiety producing drugs seem to offer possibilities (18).

SUMMARY

The design of this study was directed to testing the hypothesis that there is a relation between the concept of death and affect-producing mechanisms as demonstrated by the presentation of symbolic material which is suggestive or representative of the death concept as stimuli eliciting a GSR response.

Twenty death symbol words were selected from the literature and mixed at random with 30 noncritical words to form a word association list of 50 words. These words were presented to 40 subjects in a word association procedure, and the emotional responses of the subjects were measured by the GSR. Significant differences were obtained in the mean responses between death words and non-death

words. These results indicate that the presentation of the death symbol words elicited, at least, an unconscious affective reaction. This evidence seems to support the hypothesis being tested.

To obtain some measure of the extent to which the death concept entered into the subjects' responses, the same 20 death symbol words, which formed a part of the word association list, were used to set up a written test of interassociation. The test was aimed at obtaining an index of the tendency of the subjects to associate these words to the death concept. The results indicate that the tendency to associate these words to the death concept was significantly greater than for any other referent. Although the associative content measure is of an indirect nature, it seems that the emotional responses obtained on the GSR were actually related to the death concept.

The evidence obtained seems to demonstrate the affective nature of the death concept, and is consistent with psychoanalytic theory concerning death symbols. The data are open, however, to other interpretations than the psychoanalytic. Cautious interpretation is called for in any case in view of the indirect nature of the evidence.

REFERENCES

1. BLOCK, V. Nouveaux aspects de la methode dite "psychogalvanique" et son utilisation comme critere de tension affective. *Rev. Neurol.*, 1952, **86**, 168–171.
2. COOMBS, C. H. Adaptation of the galvanic response to auditory stimuli. *J. exp. Psychol.*, 1938, **22**, 244–268.
3. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1954.
4. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
5. IBOR, J. JUAN LOPEZ. Psychiatry and the existential crisis. In F. J. Braceland (Ed.), *Faith, reason and modern psychiatry*. New York: Kenedy, 1955.
6. JONES, H. E. The study of patterns of emotional expression. In M. L. Reymert (Ed.), *Feelings and emotions. The Mooseheart Symposium*. New York: McGraw-Hill, 1950.
7. LINDQUIST, E. F. *Statistical analysis in educational research*. New York: Houghton Mifflin, 1940.
8. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: An interim report. *J. Pers.*, 1949, **18**, 171–179.
9. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
10. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
11. MCGINNIES, E. Discussion of Howes and Solomon's note on "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 235–240.
12. MERRILL, R. M. The effect of pre-experimental and experimental anxiety on recall efficiency. *J. exp. Psychol.*, 1954, **48**, 167–172.
13. ROSANOFF, A. J. *Manual of psychiatry*. New York: Wiley & Sons, 1927.
14. SIMMEL, E. Self preservation and the death instinct. *Psychoanal. Quart.*, 1944, **13**, 160–185.
15. THORNDIKE, E. L. *The teacher's word book*. New York: Columbia Univer., 1927.
16. THOULESS, R. H. The technique of experimentation on the psychogalvanic reflex phenomenon and the phenomenon of Tarchanoff. *Brit. J. Psychol.*, 1930, **21**, 219–239; 309–321.
17. VENABLES, P. H. The relationship between PGR scores and temperature and humidity. *Quart. J. Psychol.*, 1955, **7**, 12–18.
18. WIKLER, A. The uses of drugs in psychiatric research. *Amer. J. Psychiat.*, 1956, **112**, 961–969.
19. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

Received January 29, 1957.

THE SPECIFICITY OF INTOLERANCE OF AMBIGUITY MEASURES[1]

DOUGLAS T. KENNY AND ROSE GINSBERG[2]

University of British Columbia

MUCH of the research stimulated by *The Authoritarian Personality* (1) tacitly assumes that intolerance of ambiguity is a generalized trait. At the present time, however, the basic data to substantiate or vitiate such an assumption are surprisingly not available, in spite of the plethora of measures purporting to measure intolerance of ambiguity and the fact that Frenkel-Brunswik (10, 11) expressed caution about assuming the generality of the construct when she initially introduced it. As far back as 1949 Frenkel-Brunswik, noting that the evidence for its generality was only suggestive, stated "... a much wider array of both techniques and population samples would be necessary to establish this generality with an adequate degree of definiteness" (11, p. 131). While the array of techniques for assessing intolerance of ambiguity has increased, there unfortunately has been no corresponding demonstration that the various measures intercorrelate.

The present study is concerned with examining the construct validity (3) of proposed measures of intolerance of ambiguity in terms of two procedures. The first method tests the hypothesis that a positive relationship exists between various tests presumed to measure intolerance of ambiguity. The second approach investigates the hypothesis that proposed measures of intolerance of ambiguity are positively related to authoritarianism. In light of the multiplicity of tests offered by various investigators as measures of intolerance of ambiguity, it was decided to sample a wide range of techniques. Positive intercorrelations would support the validity of the measures as indices of intolerance of ambiguity as a generalized trait. Failure to obtain significant correlations among the measures would indicate either that intolerance of ambiguity is not a unitary characteristic or that the measures are not relevant indicators of the construct.

Particular attention was given to the relevancy of the tests to the construct. In the following, a test is judged to be relevant to the construct of intolerance of ambiguity if it (*a*) elicits behavior specifically mentioned by Frenkel-Brunswik as relevant to her construct or (*b*) had been used by previous investigators as a measure of intolerance of ambiguity. With respect to the first criterion, Frenkel-Brunswik (10, 11, 12) has set forth many behavioral signs in detail, indicating that this disposition is manifested behaviorally in resistance to reversal of apparent fluctuating stimuli, the early selection and maintenance of one solution in a perceptually ambiguous situation, inability to allow for the possibility of good and bad traits in the same person, acceptance of attitude statements representing a rigid, white-black view to life, seeking for certainty, a rigid dichotomizing into fixed categories, premature closure, and remaining close to familiar characteristics of stimuli. In regard to the second criterion, a measure is regarded as having provisional relevancy if Frenkel-Brunswik or some other investigator has published a test purporting to measure intolerance of ambiguity. Of the twelve measures all are analogous to or slightly modified versions of those used by previous investigators. Three are analogous to Frenkel-Brunswik's reversible figure task (11). Detailed considerations of relevancy leading to the selection of each test are given in a later section.

In addition to the evidence provided by the intertest correlations between purported measures of intolerance of ambiguity, this study evaluates their construct validity by testing the hypothesis, derived from the theoretical network defining the construct, that intolerance of ambiguity and authoritarianism are positively correlated.

Previous research on the relation between intolerance of ambiguity and authoritarianism has produced conflicting results. Following Frenkel-Brunswik's provocative but tentative and exploratory findings (9, 10, 11, 12), a number of investigators (2, 8, 16, 18, 19, 20) have obtained evidence for a positive relation-

[1] This study was supported by a research grant, A.P. 25, from the National Research Council, Canada.

[2] Now at Indiana University.

ship. However, considerable contradictory evidence (4, 5, 6, 7, 15) has also been reported. Coulter's finding, as reported by Eysenck (8), that a questionnaire measure of intolerance of ambiguity and a test involving a series of pictures of a dog gradually changing into a cat correlate positively, is the only available evidence of a positive relationship between various tests of intolerance of ambiguity.

METHOD

Test Situations

The test situations were administered individually to the *S*s. The experimenter (R. G.) adopted a friendly and permissive approach to the *S*s who were not informed of the purpose of the study or of the true nature of the tasks.

The test battery consisted of twelve measures of intolerance of ambiguity, one attitude measure of authoritarian-submission, and seven measures of aggression. The data derived from the aggression measures will be reported elsewhere. In order to facilitate reading of the correlational data, scoring of some variables was reflected so that a high score always signifies a high degree of intolerance of ambiguity.

Reversal fluctuations. Frenkel-Brunswik (11, p. 127) has postulated that intolerance for a high rate of fluctuation in reversible figures is indicative of intolerance of ambiguity. Such tasks meet the second criterion of relevancy in that they have previously been used by Jones (13). The Necker cube, Beaunis' cubes and Schroder's staircase were presented under three conditions of instructions and three scores were obtained. Preliminary instructions, along with a practice figure, were given in order to acquaint *S* with the nature of reversals. The figures were presented for a time of one minute per figure, each figure being presented under all three conditions before proceeding to the next figure. The three instructional sets were as follows:

1. The *passive set* instructions required *S* simply to watch the figure and report each time she witnessed a change. After *S* reported the first reversal, she was asked to describe exactly what she had seen in order to avoid the continuance of any imaginary changes. The score was the number of reversals occurring in the sixty-second interval, with the score being the total of all reversals reported for all three figures. This scoring was reflected so that a high score would indicate intolerance of ambiguity.

2. Under *active set* conditions, *S* was instructed to "try to make the figure change as often as possible." The active set score was the total number of reversals during sixty seconds per figure, reported for all three figures. This variable was also reflected.

3. The *arrested set* instructions emphasized that *S* should try to stop the figure from reversing and to keep it fixed in one position. The score comprised the total reversals for all three figures, and it was reflected.

Walk's Ambiguity Intolerance Scale.[3] This attitude measure of intolerance which O'Conner (16) found to correlate positively ($r = .55$) with the California E scale consists of eight statements. An illustrative item is: "Nobody can have feelings of love and hate toward the same person." Four response categories were provided for each item, ranging from Agree, 4, to Disagree, 1, with two undecided categories, namely, Undecided: Probably Agree, 3, and Undecided: Probably Disagree, 2. A high score on the A scale is considered indicative of intolerance of ambiguity. The rationale for this scale meets both criteria of relevancy. Frenkel-Brunswik has indicated that intolerant individuals will endorse statements "... expressly designed to reveal a dichotomizing attitude, a rejection of the different, or an avoidance of ambiguities in general" (11, p. 123). The items on this scale express such dichotomizing tendencies.

Trait Discrepancy Scale.[4] According to Frenkel-Brunswik, the person who is tolerant of ambiguities will be able to tolerate and recognize the "... actual coexistence of positive and negative features in the same object..." (11, p. 115). Working on the assumption that intolerant persons would not be able to accept the coexistence of an undesirable personality trait and a socially desirable trait in the same individual, Steiner (19) devised a trait discrepancy scale that provided him with a measure of intolerance correlating positively with ethnocentrism. Using Steiner's method of scale construction, a similar scale was constructed, 25 self-concepts being rated for social desirability on a 7-point rating scale by 67 university undergraduates (28 females and 39 males). The judges were instructed to rate each trait on a continuum of social desirability and the median rating for each trait served as its social desirability scale value. The trait discrepancy scale was then constructed by pairing two sets of traits together, one of the sets having similar scale values with respect to social desirability and the other widely discrepant values. The final scale consisted of 15 such pairs.

In the test situation the scale was presented to *S* who was instructed to cross out the pair of characteristics less likely to occur together in the same person. For example, item seven was: (obedient-economical) (obedient-leisurely). The traits judged to be discrepant in social desirability scale values are "obedient-leisurely." The score was the number of discrepant traits crossed out.

Interpersonal Kind-of-Person Situation. In this test, as used by Martin (14) *S* is instructed: "Now I would like you to describe what kind of person you are." Two minutes were allowed for the task and the score was the

[3] A copy of this scale, together with the 12 attitude items measuring authoritarian submission employed in deriving the Undecided Answers measure, and the Trait Discrepancy scale has been deposited with the American Documentation Institute. Order Document No. 5462 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[4] See Footnote 3.

number of questions asked by *S* during that period. In response to any questions from *S*, the *E* replied with a set of standardized answers used by Martin. A high score is assumed to indicate intolerance of ambiguity on the grounds that "*E* confronted the *S* with an indefinite and ambiguous task and gave vague initial instructions. There were actually no correct or right responses in the situation, only a variety of alternative responses. The *S*'s behavior in regard to clarifying the ambiguity was quantified by counting the number of questions that the *S* asked before he was willing to impose his own interpretation or structure upon the task and proceed accordingly (14, p. 495)."

Interpersonal Block Situation. In this test Martin's procedure and instructions (14) were followed exactly but with some modifications of the materials. Whereas Martin presented an assortment of Vigotsky blocks to *S*, the present adaptation consists of presenting to *S* a large square of cardboard on which five design blocks from the Kohs Test and two sets of the Goldstein-Scheerer Color Form blocks are randomly assigned. *S* was instructed as follows: "I would like you to figure out what is to be done with these and then go ahead and do it." Three minutes were allowed for the task and the score was the number of questions asked during that time. As in the previous test, Martin's standard procedure for answering *S*'s questions was followed. As with the preceding test, *S* is confronted with an ambiguous situation with the possibility of many different responses; the *S* who is intolerant of ambiguity is expected to ask many questions. Martin found significant correlations between his interpersonal measures and the aniseikonic illusion.

Total Questions. The number of questions asked during the one and a half hour experimental session (excluding the questions asked during the two preceding tasks) was the score for this measure. As the whole battery of tests presented a somewhat ambiguous situation to *S* in terms of the purpose of the experiment and the nature of the experimental tasks, it was assumed that an *S* who asked many questions in this connection or posed specific questions about the procedures was attempting to structure the ambiguous situation. It will be recalled that *E* did not explain the nature or the purpose of the experiment. The *E* had merely stated that most psychological experiments necessitate the omission of such explanations. The *E*'s approach to the *S*s was permissive and nonauthoritative in order that any submissive tendencies that the intolerant individuals might have to authority figures might be overcome.

Autokinetic Effect. Block and Block (2) and Taft (20) report evidence that the intolerant individual stabilizes his judgment more quickly than others in the autokinetic experiment. Accordingly, the autokinetic situation was included in this study. Because the number of cues is minimal the autokinetic experiment would seem to present an *S* with a markedly ambiguous situation.

The apparatus used was a small container, 4¼″ in diameter and 4¼″ in height, in which a 15-watt bulb was inserted. A small aperture, about 1 mm. in size, was drilled in the side of the container and covered with four layers of thin red tissue. To heighten the autokinetic effect, a noisy motor was connected to a timer which also controlled the light, both being activated for 15 seconds during each of 10 trials.

To prepare the *S* for the autokinetic effect, detailed instructions were given prior to entering the apparatus room. In brief, the instructions emphasized that *E* would turn out the light when they entered a dimly illuminated room and that when *E* said "now" a red pin point of light would appear directly in front of the *S* at about eye level for 15 seconds. After the light went out, *S* was to estimate three things: the distance the light moved, the direction in which it moved and how confident she was about her judgment of direction. The *S* was told that she should not worry too much about the accuracy of her judgments of distance and direction, as such estimates were difficult to make under these conditions. When *S* fully understood the nature of the task, the apparatus room was entered. The *S* was seated in front of the apparatus at a distance of 12 feet. The autokinetic light source could not be seen by *S* from where she was seated, nor had the distance of this light source from the *S*'s chair been mentioned. When *S* was seated, *E* immediately darkened the room and *S* was adapted in darkness for about four minutes before the 10 trials began. During this period, *E* again reviewed the general instructions. The obtained measures were:

1. The autokinetic *mean discrepancy measure* was obtained by grouping the ten trials in successive blocks of two trials, starting from the first trial. The mean movement for each of the five blocks was calculated. If we let $i = 1, 2, 3, 4, 5$ represent the five successive blocks, an *S*'s mean discrepancy score was $X_i = (\bar{X}_1 - \bar{X}_2) + (\bar{X}_2 - \bar{X}_3) + (\bar{X}_3 - \bar{X}_4) + (\bar{X}_4 - \bar{X}_5)/4$. This variable was reflected in order that a high score would indicate intolerance of ambiguity. In keeping with Frenkel-Brunswik's (11, p. 127) assumption that intolerance of ambiguity is shown in the tendency to fix upon one of many response possibilities in an equivocal situation, persons intolerant of the unstructured autokinetic situation should be prompt in establishing an anchoring point.

2. A *consistency score* was obtained by assigning a score of one for each movement consistent with the direction of movement of the preceding trial. According to Frenkel-Brunswik's formulation (11, 12) that the

TABLE 1

CHARACTERISTICS OF THE MEASURES OF INTOLERANCE OF AMBIGUITY AND AUTHORITARIANISM

Test	Mean	*SD*	Range
Passive, Reversals	13.79	7.61	2–50
Active, Reversals	28.49	17.28	6–94
Arrested, Reversals	11.47	6.02	0–28
Walk's A Scale	17.09	3.62	8–25
Trait Discrepancy	7.37	1.95	2–12
Kind-of-Person	.88	.89	0–7
Blocks	1.67	1.27	0–6
Total Questions	3.79	3.07	0–16
Discrepancy, Autokinetic	3.41	3.98	0–26.5
Consistency, Autokinetic	3.82	2.88	0–9
Confidence, Autokinetic	42.11	7.03	22–50
Undecided Answers	5.41	3.58	0–14
Authoritarian Submission	27.54	7.71	12–43

TABLE 2
INTERRELATIONS AMONG MEASURES OF INTOLERANCE OF AMBIGUITY

Test Measures	2	3	4	5	6	7	8	9	10	11	12	13
1. Passive, Reversals	.34**	.31**	−.08	−.11	−.02	−.12	−.05	.22*	−.15	.13	−.01	.01
2. Active, Reversals		.14	.00	−.17	−.03	−.06	−.03	−.18	−.19	.01	−.08	.07
3. Arrested, Reversals			.08	.01	.02	.06	−.16	−.17	.05	−.01	.16	.20
4. Walk's A Scale				.06	−.09	.32**	.01	−.02	−.19	−.10	−.11	.57**
5. Trait Discrepancy					.17	.07	.16	−.01	.02	.06	−.08	−.07
6. Kind of Person						.19	.01	.01	−.08	−.03	−.07	.03
7. Blocks							.38**	.07	.02	.04	−.09	.07
8. Total Questions								−.20	−.01	.00	−.24*	.25**
9. Discrepancy, Autokinetic									.05	−.18	−.03	−.03
10. Consistency, Autokinetic										.17	.03	.14
11. Confidence, Autokinetic											−.26*	−.01
12. Undecided Answers												.17
13. Authoritarian-Submission												

* $p < .05$, two-tailed test.
** $p < .01$, two-tailed test.

individual intolerant of ambiguous situations tends to be rigid and consistent in his behavior when placed in ambiguous situations, the person intolerant of ambiguity should be consistent in the direction she saw the light move.

3. A *confidence estimate* was obtained for each trial, as already mentioned, in terms of very uncertain, somewhat uncertain, average amount of certainty, somewhat confident and very confident, scored from 1 to 5 respectively. The final confidence estimate score was the summation of the rating scores for all the trials. It is assumed that the intolerant person should obtain a high score, in the attempt "to deny external ambiguity as long as such denial can be maintained" (11, p. 129).

Undecided Answers. This score was derived from the answers to the eight items on the Walk scale and the 12 attitude items measuring authoritarian-submissive tendencies, to be described next. It will be recalled that the response categories to the Walk scale provided two "undecided" classifications. The authoritarian-submissive scale was similar in this respect. It is assumed that an *S* who checked either of the two undecided categories displayed tolerance for ambiguity. As the total score was the number of undecided response categories checked, this variable was reflected in the correlational study. McCandless and Holloway (15) made a similar assumption about extreme sociometric ratings. The rationale for relevancy rests upon Frenkel-Brunswik's formulation that the person intolerant of ambiguity expresses his feelings and reactions "in terms of the ends of a continuum rather than a continuum proper" (12, p. 394).

Authoritarian Submission Attitude Scale.[6] The results of the factor analysis of authoritarian items by O'Neil and Levinson (17) provided the basis for the selection of items in this measure. The scale consists of 12 items which had a factor loading of .35 or higher on their Authoritarian-Submission factor. Of the 12 items, five were from the original F scale, one was from the E scale and six were from a Traditional Family Ideology Scale. The response categories and the method of scoring were identical to those used with the Walk scale.

[6] See Footnote 3.

Subjects

The *S*s were 76 volunteer female adults. Their mean age was 23.84 years with a standard deviation of 7.47. Their education ranged from the eleventh grade to graduation from university, with an average of 13.43 years of education. Fifty-six of the *S*s were drawn from the clerical staff of the University of British Columbia. Of the remaining *S*s, 15 were teachers, 4 were students and one was a nurse.

RESULTS AND DISCUSSION

Table 1 shows the means, standard deviations and ranges of the various measures. Intercorrelations between the measures are given in Table 2. The most outstanding feature of Table 2 is the lack of evidence for positive relationship between the various measures of intolerance of ambiguity. With a two-tailed test of significance, four of the 66 correlations are significant at the .01 level of significance, and another three correlations are significant at the .05 level of confidence. Even if a one-tailed test had been employed on the grounds that the hypothesis of a positive relationship between measures of intolerance of ambiguity dictates the direction of the correlation, the main trend of the findings would not be altered. Moreover, two of the seven significant correlations obtained are not in the direction to be predicted from theory. It is also clear from Table 2 that, in the main, the purported measures of intolerance of ambiguity do not correlate significantly with the authoritarian-submission scale, indicating that the hypothesis of a positive relationship between authoritarian-submission and intolerance of ambiguity is not sustained.

Only two of the 12 correlations employed to test this hypothesis are significant.

In so far as the present measures may be regarded as relevant indicators of the intolerance of ambiguity construct, the results offer little support for a general construct of intolerance of ambiguity. It would be unwarranted to infer that a person who is intolerant in one situation will be intolerant in all equivocal situations. Moreover, it would be equally unwarranted to assume that intolerance of ambiguity is invariably linked with authoritarianism. These conclusions do not mean, however, that the construct should necessarily be discarded. It may, rather, be less general or broad in scope than had been initially assumed. Future research may discover a number of distinct or relatively independent dimensions of intolerance of ambiguity rather than just one unique generalized factor.

Another inference from the data is that the nature of the task is a crucial element in the discrepant results of studies in this area. If no common factor of intolerance of ambiguity runs through all the tests employed in its name, then the discordant results are not too surprising. This conjecture seems quite plausible in light of our failure to find any such unique factor.

SUMMARY

The major focus of the present study was a test of the construct validity of proposed measures of intolerance of ambiguity. Seventy-six female *Ss* were given a battery of tests purporting to measure intolerance of ambiguity and authoritarian-submission. Only seven of the 66 correlations among measures of intolerance of ambiguity were significant at the .05 level, two of these having a relationship opposite to those predicted. In the main, these measures did not correlate significantly with a scale of authoritarian-submission.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BLOCK, J., & BLOCK, JEANNE. Intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1950, **19**, 303–311.
3. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.
4. DAVIDS, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415–420.
5. DAVIDS, A. The influence of ego-involvement on relations between authoritarianism and intolerance of ambiguity. *J. consult. Psychol.*, 1956, **20**, 179–184.
6. DAVIDS, A., & ERIKSEN, C. W. Some social and cultural factors determining relations between authoritarianism and measures of neuroticism. *J. consult. Psychol.*, 1957, **21**, 155–159.
7. ERIKSEN, C. W., & EISENSTEIN, D. Personality rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386–391.
8. EYSENCK, H. J. *The psychology of politics*. London: Routledge & Kegan Paul, 1954.
9. FRENKEL-BRUNSWIK, ELSE. Tolerance toward ambiguity as a personality variable. *Amer. Psychologist.*, 1948, **3**, 268. (Abstract)
10. FRENKEL-BRUNSWIK, ELSE. A study of prejudice in children. *Hum. Relat.*, 1948, **1**, 295–306.
11. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual variable. *J. Pers.*, 1949, **18**, 108–143.
12. FRENKEL-BRUNSWIK, ELSE. Personality theory and perception. In R. R. Blake and G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 356–419.
13. JONES, M. B. Authoritarianism and intolerance of fluctuation. *J. abnorm. soc. Psychol.*, 1956, **50**, 125–126.
14. MARTIN, B. Intolerance of ambiguity in interpersonal and perceptual behavior. *J. Pers.*, 1954, 494–503.
15. MCCANDLESS, B. R., & HOLLOWAY, H. D. Race prejudice and intolerance of ambiguity in children. *J. abnorm. soc. Psychol.*, 1955, **51**, 692–693.
16. O'Conner, P. Ethnocentrism, "intolerance of ambiguity," and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526–530.
17. O'NEIL, W. M., & LEVINSON, D. J. A factorial exploration of authoritarianism and some of its ideological concomitants. *J. Pers.*, 1954, **22**, 449–463.
18. SIEGEL, S. Certain determinants and correlates of authoritarianism. *Genet. Psychol. Monogr.*, 1954, **49**, 187–229.
19. STEINER, I. D. Ethnocentrism and tolerance of trait "inconsistency." *J. abnorm. soc. Psychol*, 1954, **49**, 349–354.
20. TAFT, R. Intolerance of ambuiguity and ethnocentrism. *J. consult. Psychol.*, 1956, **20**, 153–154.

Received May 28, 1957.

PERCEPTUAL REACTIVITY TO THREAT AND NEUTRAL WORDS UNDER VARYING EXPERIMENTAL CONDITIONS[1]

EMORY L. COWEN AND PAUL A. OBRIST

University of Rochester

THE purposes of the present research are twofold: (*a*) to investigate the effect of systematic variation of expectancy on perceptual report thresholds for threat and neutral words and (*b*) to provide an experimental test of the usefulness of "hypothesis theory" in explaining the elevation of thresholds for report of threat words observed in some previous investigations.

In an early controversial study, McGinnies (14) found significantly higher thresholds for report of taboo words when compared to neutral ones. That these data offer support for the existence of some organismic "defense" process, has been challenged by Solomon and Howes (17), among others, who argue that McGinnies failed to control for differential familiarity for the two classes of word stimuli, as well as for deliberate or conscious inhibition of report. Subsequent investigations including those of Eriksen and Brown (4), Carpenter, Wiener, and Carpenter (1), and Cowen and Beier (2) in the area of perception, as well as studies by Lazarus, Eriksen, and Fonda (11) and Lazarus and Longo (12), in which personality attributes or clinical typologies have been related to differential reaction to emotional and nonemotional stimuli, has led to a reaffirmation (3) of the concept of "perceptual defense."

On the other hand, another line of research has developed as exemplified by the work of Freeman (5, 6), Lacy, Lewinger, and Adamson (9), Postman, Bronson, and Gropper (16), and others, which has moved in the direction of rejecting this theoretical position. Instead it has been proposed that the phenotypical phenomena frequently cited as illustrations of "perceptual defense" may be more parsimoniously viewed as special instances of already established perceptual principles such as set, or hypothesis behavior. Illustratively, in the experimental study by Postman *et al.* (16) it was found that by controlling for the differential word frequency of taboo and neutral words, response thresholds for the same taboo words used earlier in the McGinnies study (14) were now significantly *lower* than those for frequency-matched neutral words. Elsewhere Postman proposes:

> What appears to be perceptual defense results from the dominance of strong alternative hypotheses rather than from active repression of the inimical or dangerous. In the presence of partial information, strong hypotheses incompatible with the threatening stimulus may be evoked. . . . If this is the case the subject will appear to be defending himself against perception. If, however, hypotheses related to the negative stimuli are strong, the opposite of defense will appear to operate (15, p. 300).

In the present research, strength of hypothesis for threat and neutral words has been varied systematically. From Postman's theoretical position it would be predicted (*a*) that as expectancy or set for threat words is increased experimentally, the discrepancy in threshold of report for threat and neutral words will decrease, and (*b*) that as 100 per cent expectancy is reached for threat words, thresholds of report for threat words will become (significantly) lower than those for neutral words.

METHOD

Thirty male and thirty female introductory psychology student volunteers served as *S*s in the present research. These *S*s were randomly assigned by sex to one of three experimental groups, each of which contained ten males and ten females.

All *S*s were initially informed that they were to decipher words from a series of booklets consisting of thirty carbon copies of a single five-lettered word, electrically typed in capital letters, and arranged in order from the most blurred copy to the clearest. This technique is described more fully elsewhere (2). Following two practice booklets the three experimental preconditions were initiated. A list of 36 words was read to each *S* with the instruction that the final test series would contain some words similar to those on the reading list. There were three groups, as follows:

Group I (0 per cent threat expectancy) was read only neutral words.

Group II (33 per cent threat expectancy) was read a list containing 12 threat words and 24 neutral words.

Group III (100 per cent threat expectancy) was read all threat words.

[1] Portions of the present paper were presented to Division 8 at the Amer. Psychol. Ass. meetings in Chicago, August 1956.

TABLE 1
PAIRED THREAT AND NEUTRAL WORDS WITH FREQUENCIES

Threat	Neutral	Frequency[a]
Bitch	*Juror*	4
Penis	*Rayon*	1
Whore	*Derby*	3
Death	*Clear*	100+
Urine	*Humid*	1
Cheat	*Rural*	18–19
Stink	*Pivot*	4
Flunk	*Sedan*	2

[a] Based on general frequency of occurrence in one million words (18).

The threatening or neutral nature of the words for the reading list (as well as the final test series) had previously been established empirically by means of ratings by graduate students in psychology, with respect to the dimension of stimulus threat value. Only those words with consistent interjudge agreement in designation as either threat or neutral words were included in the reading lists and final test series. For all groups, *Ss* were asked to repeat each word after the *E*. This precaution was taken not only to make sure that the *S* had heard the word correctly, but also to give *Ss* experience in the actual verbalization of threatening stimuli.

All *Ss* were then given the test booklets, in random order, to decipher. This final test series, individually administered, consisted of 16 words, eight threat and eight neutral, *none* of which had appeared on any of the earlier reading lists. The threat words contained four of the "sexually-taboo" variety, and four others that did not involve any sexual taboo. Each of the eight threat words had been matched with a neutral word on the basis of identical tabled frequencies according to the Thorndike-Lorge word frequency count (18). Where no word frequency was given for a threat word, an arbitrary frequency of one was assigned, and matching of the neutral word proceeded on this basis. Table 1 presents the test series of words and their frequencies.

The *Ss* were allowed two seconds per page after which they were required to turn to the next page. The *E* recorded total trials required for correct report of each word. Guessing was encouraged, and all prerecognition guesses were recorded. At the end of the perception experiment, all *Ss* were asked to write down as many words as they could recall from the original pre-experimental list of words read to them. They were given three minutes for this task.

RESULTS AND DISCUSSION

Report thresholds. Mean thresholds for threat and neutral words as well as the departure of the mean D score between these two classes of words from zero are presented for each group in Table 2. Table 3 summarizes a three-way analysis of variance for perceptual report thresholds, the main effects of which include type of words (threat vs. neutral), sex, and experimental conditions (level of set). Since all *Ss* were given both classes of word stimuli, this effect is regarded as a within subject effect. Accordingly, a mixed design Type III, as described by Lindquist (13), was employed as the basic model. Relative to the purposes of the present experiment, the most salient results in this table are the significant words-by-conditions interaction and the significant main effect of words.

The significant words-by-conditions interaction indicates that there are reliable differences in threat minus neutral word thresholds (D scores) for the three experimental conditions. From inspection of the last column of

TABLE 2
MEAN PERCEPTUAL REPORT THRESHOLDS AND D SCORES FOR EXPERIMENTAL CONDITIONS AND SEX

Condition	Sex	Threat	Neutral	D score[a]	Sig. of D score from 0
I (0% threat)	Male	22.15	20.79	+1.36	.05–.1**
	Female	22.15	18.84	+3.31	.01
	Total	22.15	19.82	+2.33	.01
II (33% threat)	Male	21.18	19.93	+1.25	.1–.2
	Female	22.91	20.45	+2.46	.01
	Total	22.04	20.19	+1.86	.01
III (100% threat)	Male	21.21	21.59	−0.38	.5–.6
	Female	21.93	21.08	+0.85	.05–.1
	Total	21.57	21.34	+0.24	.5–.6

[a] Plus (+) signs indicate higher thresholds for threat words.
** p value tabled on basis of two-tailed test.

TABLE 3
ANALYSIS OF VARIANCE FOR PERCEPTUAL REPORT THRESHOLDS
(Lindquist Type III Design)

Source	*df*	Mean Squares	*F*	*P*
Between *Ss*	(59)			
Sex	1	0.21	—	—
Conditions	2	2.33	—	—
Sex × Conditions	2	11.10	—	—
Error (between)	54	38.42		
Within *Ss*	(60)			
Words	1	65.50	28.48	.001
Words × Sex	1	16.05	6.98	.01–.05
Words × Conditions	2	12.11	5.27	.01
Words × Conditions × Sex	2	0.34	—	—
Error (within)	54	2.30		
Total	119			

Table 2, it can be seen that there is a drop in discrepancy between threat and neutral word thresholds, as we strengthen experimentally the set for threat stimuli. Table 2 also suggests that a major component of this drop may be the increase in neutral word thresholds. However, analysis of variance comparing the mean thresholds for neutral words across conditions failed to demonstrate significant differences. We may note that direct analysis of neutral word thresholds compares different *groups* of *Ss*, whereas D-score analysis is based on each *S* serving as his own control. In the former instance we thus introduce intragroup variability (error variance) deriving from such sources as differential visual acuity, inconstancies in testing conditions etc., and consequently reduce the likelihood of demonstrating significant neutral word threshold differences. Accordingly, we view the earlier noted decrease in threat-neutral discrepancy from Condition I to Condition III as confirming the first prediction from hypothesis theory, consistent with the earlier findings of Postman *et al.* (16).

The significant main effect of words in Table 3, viewed conjunctively with the actual group means presented in Table 2, indicates that threat words on the whole have significantly higher thresholds than neutral ones. Although it is probable, on the basis of the words-by-conditions interaction, that the several conditions contribute differentially to this main effect, these data are nevertheless in sharp contrast to the empirical findings of Postman *et al.* (16). The latter investigators report that after matching threat and neutral words for word frequency, *lower* report thresholds were found for threat words under *all* of their experimental conditions. Postman *et al.* (16) attribute their results to the probable underestimation of functional word frequencies for threat words in standard word frequency tables. Notwithstanding the rather precise control for word frequency in the present experiment, the Postman *et al.* (16) findings are not duplicated. In fact, with respect to this segment of the analysis, the present results go significantly in the opposite direction.

In addition to this discrepancy at the empirical level it is important to note that the overall higher threat thresholds observed in the present study negate the second prediction deriving from hypothesis theory (i.e., that threat words should be seen significantly more rapidly than neutral ones where a 100 per cent expectancy for the former has been established). Under none of the present experimental conditions, including 100 per cent threat expectancy, are threat words seen more rapidly than neutral ones. In fact, for the 100 per cent threat condition, the overall trend remains in the direction of higher thresholds for threat words, albeit not significantly so.

To this point, the discussion has been limited to the derivations from hypothesis theory, which we have set out to study experimentally. Our findings, however, have some implications for the word-frequency control issue originally developed by Solomon and Howes (17). For the specific experimental conditions and word stimuli used in this study, it is clear that an explanation of the elevation of threat word thresholds in terms of differential word frequency would not be defensible. This is not to say that word frequency cannot or has not constituted an uncontrolled element spuriously influencing thresholds in other experiments. What remains, however, is the fact that threat word thresholds continue to be elevated in a situation where word frequency is either ruled out as a contaminating influence, or if not, is as Postman *et al.* (16) suggest, operating in a direction opposite to the findings (i.e. to lower thresholds for threat stimuli by underestimation of functional usage of threat words).

The present investigators were nevertheless somewhat perplexed in trying to develop a frame of reference in which the discrepancies between our findings and those of Postman *et al.* (16) with respect to the relative elevation or depression of threat word thresholds might be resolved. To us it seemed that a promising lead might be contained in Lazarus' earlier critique (10) of the Postman *et al.* study, in which it was proposed that tabled *threat* word frequencies do not necessarily correspond with their general familiarity to college students. Lazarus further suggested that "... findings in terms of mean differences in thresholds for threat and neutral words could be made to go in either direction by the appropriate selection of words on the basis of such variables as context, contrast, sequential probabilities, and a host of other factors not directly related to word frequency" (10, p. 397).

As we examined the actual word lists utilized in the two investigations it appeared probable to us, quite apart from the threat word issue raised by Lazarus, that many of the *neutral* words from our study had substantially higher functional usage for college students than did neutral words with comparable tabled frequencies, used by Postman *et al.* (16). A small additional study was undertaken in order to subject this hunch to experimental test.

Ten "neutral" words were selected, five from the Postman *et al.* series and five from our own. These words were chosen largely in terms of our preconceptions regarding their functional frequencies. The words were placed in a fixed random order, and given to 30 judges, all of whom were either staff members or gradu-

TABLE 4

NEUTRAL WORDS FROM TWO EXPERIMENTS, ORDER OF PRESENTATION, FREQUENCY, AND JUDGES' RANKS

Words	Order	Frequency	Mean rank	Rank of $\bar{x}$ Rank
Postman *et al.*				
Lathe	2	2[a]	5.90	6
Lotus	4	2	8.23	8
Tiara	7	1	8.30	9
Capon	8	1	8.93	10
Clove	9	8	6.83	7
Mean		2.8	7.66	8
Cowen & Obrist				
Rayon	1	1	2.53	2
Humid	3	1	1.17	1
Juror	4	4	3.50	3
Pivot	6	4	4.33	4
Derby	10	3	5.17	5
Mean		2.6	3.34	3

[a] Based on Thorndike-Lorge (18) general count for frequency of occurrence in one million words.

ate students in the psychology department at the University of Rochester. The judges were asked to rank order the 10 words on the basis of estimated frequency of usage by a mixed (male and female) college population. Mean functional frequency ranks were then computed for all words.

Table 4 presents the two sets of words, together with their order of presentation, frequency of occurrence in one million words according to the Thorndike-Lorge general count (18), and mean judge rankings. We note first that the tabled frequencies for the two sets of words are highly comparable. Were there some correspondence between tabled frequencies and estimated functional usage, it might be anticipated that judges' rankings for the latter dimension would be essentially random across the two sets of words. Inspection of Table 4 indicates that this is far from the case. Estimated functional frequency ranks for our words, run from one to five. There is, in fact, not a single "misplacement" of ranks for the words from the two studies. This global impression is visible as well in the analysis of individual word rankings. Three of the Postman *et al.* words, *tiara*, *lotus*, and *capon* have 75 per cent or more of their ranks at eight, nine, and ten. On the other hand, our word *humid* has 25 "one" ranks and five twos. While we are willing to concede that this specific ranking may, in part, reflect some climatic preoccupation of Rochesterians, the basic trend of differences for the functional frequencies of the two sets of words seems highly reliable.

The foregoing demonstration seems to establish, at least within the range of frequencies studied, the possibility of gross disparity between the tabled values characteristically used in selecting words for "perceptual defense" studies, and estimated functional usage of words. Moreover, the particular frequency band examined is of some importance, since, due to the usually low tabled frequencies for threat words, neutral control words are most often selected from within this range.

The demonstrated discrepancy between tabled frequencies and estimated functional frequencies is alarming! For not only must we face up to past speculations regarding the probable systematic underestimation of threat word frequencies (10), but we must now add to this the discomforting recognition that within the critical low frequency range there may be no correspondence between estimates of word frequency derived from adult and juvenile books, newspapers and magazines on the one hand and the frequencies inherent in the every day language of the college student. Since "perceptual defense" studies have been executed almost unilaterally on the college sophomore, or some variant thereof, it would seem that the reference population for determining word frequencies should most appropriately be such a group in future work. This will require the empirical establishment of word frequency norms for college students (or whatever other reference group is to be studied perceptual defense-wise). The technique we have reported here does no more than to approximate this end very crudely.

One further disconcerting comment may be appropriate. When a "perceptual defense" study, in which word frequency is to be controlled is undertaken, the investigator characteristically matches neutral word frequencies to available tabled frequencies for threat words. In so doing, he has a very substantial pool of low frequency words to draw from, which, on the basis of present findings, may not necessarily correspond to actual usage frequencies. In such a situation there may develop the unintentional but nevertheless quite serious danger of systematic error in the selection of specific words from a group with identical frequencies, on the basis of unspecified biases and predilections. To the extent that, within this key low frequency range, tabled and functional frequencies do not correspond, we are faced with an important problem of representative and systematic design (7). For if we are ultimately interested in functional frequencies, and n sets of equated tabled frequencies represent n continua of functional frequencies, our fundamental N with respect to functional frequency is one, and generalizations about word frequency may be restricted to the (unknown) functional frequencies of the *specific* set of neutral control words employed.

It is quite possible that the reversal in findings of the two studies with respect to relative elevation or depression of threat words thresholds, is in part explainable in terms of differential functional frequencies. However, even though we have demonstrated that some of our own neutral words have higher functional frequencies than some of those with comparable tabled frequencies used by Postman *et al.*, there are no data from either study contrasting the functional frequencies of threat and neutral words. We can here do no more than speculate along with Lazarus (10) that for the pertinent reference population (i.e. college students) the threat words used may have higher functional frequencies than "matched" neutral control words (for both studies). On the basis of such speculation, we believe that our earlier statements concerning elevation of threat word thresholds may be defensible. At the same time, it would be unrealistic to fail to admit that, at least within the critical low frequency range, frequency "matchings" in studies to date may be more a matter of fantasy than fact. The necessary *desideratum* must be the determination of functional frequencies

for both classes of word stimuli. To this may be added the ultimate goal of ascertaining individual frequency differences in word usage (10). These ends may only be achieved empirically.

Sex differences. Returning to Table 3, we note a significant words-by-sex interaction, indicating that there is a difference between sexes in the discrepancy between threat word and neutral word thresholds. Table 2 suggests that this difference is due to the consistent tendency for males to have lower report thresholds than females for threat stimuli. These data are in direct support of the earlier comparable finding by Postman *et al.* (16). Of the several possible interpretations offered by the latter investigators to explain this phenomenon, that of possible differential readiness to report threat words seems best to accomodate the data.

Taboo vs. non-taboo threat words. In order to check further on the possible operation of conscious inhibition, the threat words were divided into taboo threat (bitch, penis, whore, and urine) and non-taboo threat words (death, cheat, stink, flunk). Again, using a Lindquist Type III analysis (13), mean D scores for the four taboo pairs vs. the non-taboo threat words were compared. No significant differences were observed in the D scores for the two types of words, although there was some tendency for the taboo words to have higher D scores from their matched neutral words, than did the non-taboo threat words from theirs ($p = .20$). It may be pertinent to note that *Ss* were given experience in verbalizing threat words overtly in Groups II and III during the reading period so as to reduce the possibility of deliberate inhibition. Perhaps the reduction in conscious inhibition deriving from this procedure accounts for our failure to find significant differences between taboo and non-taboo threat thresholds in the present investigation.

Prerecognition guesses. A three-way analysis of variance (Type III) was also computed for number of prerecognition guesses. Although these findings are not tabled the results are quite clear-cut. There are two significant effects observable, the main effects of words ($p < .01$) and conditions ($.01 < p < .05$). Inspection of means indicates that more guessing has occurred for threat words than for neutral ones, and that guessing behavior is highest for Group II (33 per cent threat). The former finding, although partially confounded by the greater number of threat word trials, seems to reflect at least some element of genuine differential guessing behavior for the two classes of word stimuli. Illustratively, in all six subgroups, including males in Group III (where threat word threshold is actually slightly lower than neutral word threshold) there are significantly more prerecognition guesses for the threat words.

The increase in guessing behavior noted in Group II is highly reliable for both sexes and for both types of words. Here we might speculate that the crucial determinant is the ambiguity of the initial reading list, which was the only one of the three to contain words from two clearly differentiable *genera*.

Recall behavior. A simple two-way, conditions-by-sex analysis of variance was computed for total number of words from the pre-experimental reading list recalled by *Ss* during the postexperimental period. A highly significant conditions effect ($p < .001$) was observed here, with markedly superior recall by Group III (100 per cent threat). These data are explainable in terms of the relatively finite reference population from which the threat words were drawn.

Implications. In relating our findings to current theoretical controversies, it is apparent that the principle of set or perceptual hypothesis behavior is a sufficient basis for accommodating the observed decrease, from Condition I to Condition III, in threat-minus-neutral discrepancies. On the other hand, the elevation of threat thresholds across the board, in contrast with those for neutral thresholds, constitutes a datum not readily explainable within the framework of hypothesis theory as stated by Postman (15). If we proceed from the assumption that the present neutral control[1] words constitute an appropriate matching for the threat stimuli, then the latter finding would be more readily interpretable by a concept of response interference from threatening stimuli, or differential response hierarchy along the lines proposed by Eriksen (3), Eriksen and Browne (4), Hochberg, Haber and Ryan (8), and others. In that case, it would not only be feasible but necessary to view these two presumably antagonistic theoretical explanations as compli-

mentary in order to explain adequately all of the findings of the present study.

If, however, our supplementary findings suggesting a lack of correspondence between tabled and functional frequencies are considered, we must recognize the strong possibility that earlier tests of the relationship between word-type (threat and neutral) and response threshold have been inappropriate.

SUMMARY

In order to test two predictions deriving from hypothesis theory, sixty *S*s divided into three groups were tested for perceptual reactivity to threat and neutral words matched for word frequency. The experimental groups differed only in terms of induced strength of expectation for threat words from 0 per cent to 100 per cent.

The findings support the prediction that as strength of set for threat stimuli increases, the discrepancy between threat and neutral word thresholds will decrease. But contrary to predictions from hypothesis theory, threat words were found to have higher report thresholds than neutral words in all conditions, even with word familiarity controlled.

A supplementary study, however, demonstrated gross disparities between tabled frequencies and estimated functional frequencies of neutral words, giving rise to several fundamental methodological considerations regarding "perceptual defense" research.

Several empirical analyses dealing with prerecognition guessing behavior, and postexperimental recall behavior are reported.

REFERENCES

1. CARPENTER, B., WIENER, M., & CARPENTER, JANETH T. Predictability of perceptual defense behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 380–383.
2. COWEN, E. L., & BEIER, E. G. Threat-expectancy, word frequencies, and perceptual prerecognition hypotheses. *J. abnorm. soc. Psychol.*, 1954, **49**, 178–182.
3. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175–182.
4. ERIKSEN, C. W., & BROWNE, C. T. Experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, **52**, 224–230.
5. FREEMAN, J. T. Set or perceptual defense? *J. exp. Psychol.*, 1954, **48**, 283–288.
6. FREEMAN, J. T. Set versus perceptual defense: A confirmation. *J. abnorm. soc. Psychol.*, 1955, **51**, 710–712.
7. HAMMOND, K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, **51**, 150–159.
8. HOCHBERG, J. E., HABER, S. L., & RYAN, T. A. "Perceptual defense" as an interference phenomenon. *Percept. Motor Skills*, 1955, **5**, 15–17.
9. LACY, O. W., LEWINGER, N., & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.
10. LAZARUS, R. S. Is there a mechanism of perceptual defense? A reply to Postman, Bronson, and Gropper. *J. abnorm. soc. Psychol.*, 1954, **49**, 396–398.
11. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–483.
12. LAZARUS, R. S., & LONGO, N. The consistency of psychological defenses against threat. *J. abnorm. soc. Psychol.*, 1953, **48**, 495–499.
13. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
14. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
15. POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.*, 1953, **60**, 298–306.
16. POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–225.
17. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–271.
18. THORNDIKE, E. L., & LORGE, I. *The teachers' word book of 30,000 words*. New York: Teachers Coll., Columbia Univer., 1944.

Received May 2, 1957.

AUTHORITARIANISM, PUNITIVENESS, AND PERCEIVED SOCIAL STATUS[1]

ALAN H. ROBERTS AND RICHARD JESSOR[2]

University of Colorado

THE problem dealt with in this study is the effect of perceived social status upon the relationship between authoritarianism and punitiveness. Previous studies concerned with this problem (1, 2, 3, 4, 5, 8) have tended to yield somewhat inconsistent results. These inconsistencies may perhaps be resolved in part if the social cues present in the interpersonal situation are taken into account, the relationship between authoritarianism and punitiveness being conceived to be mediated by social stimulus variables whose meaning derives largely from past learnings. Specifically, high authoritarians may be assumed to have learned to avoid expressing overt and direct aggressiveness towards individuals perceived to be of higher status than themselves and instead, to displace this aggressiveness to out-group members, violators of conventional values, and low-status figures in general. Theoretical support for this assumption may be inferred from the descriptions of "authoritarian submission" and "authoritarian aggression" in *The Authoritarian Personality* (1, pp. 230–234).

Four specific hypotheses were investigated.

1. When the perceived status of the frustrators is not considered in the analysis, there is no difference between high and low authoritarians with respect to punitiveness.

2. High authoritarians, more often than low authoritarians, tend to respond with direct hostility and aggressiveness when frustrated by low status figures.

3. High authoritarians, more often than low authoritarians, tend to respond with indirect, displaced hostility and aggressiveness when frustrated by high status figures.

4. With respect to punitiveness, low authoritarians manifest no difference in reaction to high and low status frustrators.

METHOD

Subjects

A scale[3] consisting of 28 items from the California F (authoritarianism) scale (1) was administered to approximately 800 students in an introductory class in psychology at the University of Colorado. From these 800 respondents, the 60 individuals scoring highest and the 60 individuals scoring lowest on the F scale were asked to volunteer for this study. Of these, 48 high scorers and 53 low scorers participated in the experiment. The two groups of *S*s were made up of approximately equal numbers of males and females, and the median age of *S*s in both groups was 20. The mean F-scale score of the High group was 96.75; the *SD* was 5.80. The mean F-scale score of the Low group was 47.00; the *SD* was 5.02.

Stimulus Materials

A series of 24 separate stimulus cards was constructed, each card portraying two individuals, one of whom is shown to be frustrating the other.[4] The cards are very similar in appearance to the *Rosensweig Picture-Frustration Study* (6), and may be considered as a revision of that technique for our specific purposes. On twelve of the cards in the present study (the "high status cards"), the frustrator is depicted as a person of high social status with respect to the person being frustrated. On the remaining twelve cards (the "low status cards"), the frustrator is depicted as a person of low social status with respect to the person being frustrated. The social status of the stimulus figures was intended to vary along the dimensions of age, education, income, occupation, and recognizable stereotype of authority in our society. Cues appropriate to these dimensions were manipulated in the drawings.

[1] This study was supported in part by financial aid from the psychology department of the University of Colorado and from the Office of Naval Research grant to J. R. Gibb.

[2] The first author is now at Elgin State Hospital, Elgin, Illinois. The second author was a Social Science Research Council fellow at the time this article was written.

[3] The scale was of the usual Likert type, but the number of response categories was reduced from six to four to permit machine scoring of the completed questionnaires. Each item was scored 1, 2, 4, or 5 according to whether the *S* disagreed very much, disagreed a little, agreed a little, or agreed very much, respectively. If an item was not answered it received a score of 3.

[4] Reproductions of the 24 stimulus cards used in this study have been deposited with the American Documentation Institute. Order Document No. 5460 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $2.25 for microfilm or $5.00 for photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

An example of a high status card is Card 10 (Army Officer to Enlisted Soldier: "I know I promised you a pass this weekend, but you'll have to stay here and stand guard duty"). An example of a low status card is Card 11 (Secretary to Boss: "I'm too busy to type that now").

To determine how successfully the social cues were manipulated in the drawings, the face validity of the stimulus cards was ascertained. After the main part of the experiment was concluded, all *Ss* were asked, as part of a "test of perceptual sensitivity," to guess the age, occupation, yearly salary, and years of education of both figures on six of the high and six of the low status cards, exposed for 30 seconds each. These data clearly indicate that the relative status relationships were generally perceived in the intended direction by a large majority of the *Ss*.[5] There were no differences between high and low authoritarians in these judgments.

Administration and Scoring

The *Ss* were tested in two intermixed groups of approximately 50 *Ss* each. By means of an opaque projector, the stimulus cards were shown separately on a large screen for 60 seconds, and *Ss* were instructed to write in their answer books the very first reply that came to mind in answer to the comment of the frustrating figure.

Four scoring variables[6] to characterize the responses of the *Ss* in terms of punitiveness were defined in detail and denoted by specific examples. The definitions correspond closely to those of Rosenzweig (7). Very briefly, the four variables were:

1. Personal Extrapunitiveness (P)—the response is directly hostile or aggressive toward the frustrator.

2. Nonpersonal Extrapunitiveness (N)—the hostile or aggressive response is displaced or directed toward some person, object, or event other than the frustrator.

The sum of the two foregoing scores was termed Total Extrapunitiveness (E).

3. Intrapunitiveness (I)—blame or censure is directed by *S* upon himself.

4. Impunitiveness (M)—the frustrating situation is construed not to be frustrating or to be beneficial.

Interscorer reliability was assessed by having three psychologists[7] independently score the same 15 protocols. For 360 scored responses, agreement between A and B was 77%, between A and C was 71%, and between B and C was 94%. All protocols were then scored by the first author without knowledge of the *Ss*' F-scale score. The total frequency of P, N, I and M scores was computed for each protocol, and the E score was obtained by summing the P and N scores.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF FIVE PUNITIVENESS CATEGORIES FOR THE TOTAL TEST AND FOR HIGH AND LOW STATUS CARDS SEPARATELY

Measure	High F (N = 48)		Low F (N = 53)		t
	Mean	*SD*	Mean	*SD*	
Total Test:					
E	18.06	2.37	17.57	2.69	.95
P	5.83	2.34	5.14	2.85	1.32
N	12.23	2.29	12.43	2.75	.39
I	1.67	.94	2.00	1.26	1.60
M	4.25	2.19	4.34	2.24	.29
High Status Cards:					
E	9.48	1.47	9.67	1.78	.42
P	.67	1.00	2.75	2.02	6.61**
N	8.81	1.40	6.92	2.08	5.33**
I	.46	.58	.53	.72	.55
M	2.04	1.40	1.77	1.48	.94
Low Status Cards:					
E	8.59	1.46	7.91	1.48	2.25*
P	5.17	1.83	2.39	1.43	8.37**
N	3.42	1.57	5.52	1.40	7.01**
I	1.21	.66	1.47	.80	1.82
M	2.21	1.27	2.57	1.31	1.39

* Significant beyond the .05 level of confidence (two-tailed test).

** Significant beyond the .01 level of confidence (two-tailed test).

RESULTS AND DISCUSSION

The significance of the mean differences between high and low authoritarians with respect to E, I, M, N, and P on all 24 stimulus cards together, and on the 12 high status and 12 low status cards separately, was tested by means of the t test for independent large samples. The results of these comparisons may be seen in Table 1.[8]

With respect to Hypothesis 1, no significant differences in E, I, M, N, or P scores on the test *as a whole* could be found. While negative findings cannot verify a hypothesis, this lack of difference is fully consistent with our first hypothesis.

When the high and low status cards are analyzed separately, the differences between

[5] For example, on Card 3, the stimulus figure depicted as having the higher status was rated to be younger than the low status figure by only 4 of the 101 *Ss*, was rated to have less education by 1 of the 101 *Ss*, and was rated to have less income by none of the *Ss*. The high status figure was perceived as having a business or professional occupation by 96 *Ss*, a semiprofessional occupation by 4 *Ss*, and was labeled a "gentleman" by one *S*. Similarly, the low status figure was perceived as a waitress by 98 *Ss*, a maid by 2 *Ss*, and a housewife by the remaining *S*. A complete summary table of the face validity data has been deposited with ADI, cf. Footnote 4.

[6] Complete definitions of the scoring variables have been deposited with ADI, cf. Footnote 4.

[7] The authors are indebted to James Selkin for his assistance in establishing the interscorer reliability of the punitiveness categories.

[8] All raw score data have been deposited with ADI, cf. Footnote 4.

the High and Low F groups with respect to P and N scores are all significant in the predicted direction beyond the .01 level of confidence. We may therefore reject the null statement of Hypotheses 2 and 3. It appears clear that high authoritarians, more often than low authoritarians, tend to respond with personal hostility to low status figures and with displaced hostility to high status figures. Conversely, high authoritarians, less often than low authoritarians, tend to respond with personal hostility to high status figures and with displaced hostility to low status figures. These results suggest that findings of a relationship between total extrapunitiveness and authoritarianism (2, 5, 8) may be at least partly accounted for in terms of this previously overlooked status variable rather than by assuming greater generalized extrapunitiveness on the part of high authoritarians. This interpretation is further supported by the finding in Table 1 of a significant difference between high and low authoritarians with respect to total E on the low status cards taken alone. It is only on these cards that high authoritarians are more extrapunitive than low authoritarians.

The final hypothesis was to the effect that while the responses of high authoritarians to frustration are, at least in part, mediated by the perceived status of the frustrator, low authoritarians tend to respond to frustration relatively independently of the status of the frustrator.

In order to test Hypothesis 4, the difference between the mean P score on the high status cards and on the low status cards was statistically evaluated for the High and for the Low F groups separately. In the High F group this difference was clearly significant beyond the .01 confidence level (t, for related measures, of 17.31). The same analysis for the Low F group showed the difference to be insignificant. These data, therefore, strongly support Hypothesis 4; high authoritarians respond with direct personal hostility to low status figures significantly more often than they do to high status figures. Low authoritarians appear to respond with direct personal hostility independently of the status of the frustrator.

Hypothesis 4 includes a prediction about displaced or indirect hostility as well as about direct aggression. Hence, the same kind of analysis as described for P scores was made for the N score. This analysis shows that *both* high and low authoritarians tend to respond with displaced hostility to high status figures more often than they do to low status figures, both cases significant beyond the .01 level of confidence. This result does not accord with Hypothesis 4, but it is somewhat tempered by the fact that the difference for the High F group (5.40) is significantly greater than the difference for the Low F group (1.40), significant beyond the .01 level of confidence. In short, the data tend to support Hypothesis 4, although one finding does require certain reservations.

Several limitations of this study must be noted. In the first place, the experimental groups included only the most extreme segments of an initial group of 800 *Ss*. Generalization to the larger group of less extreme cases is not warranted on the basis of this study alone. Second, the semiprojective test used in this research was not fully refined. Some of the stimulus cards proved to be undiscriminating, and the technique as a whole would undoubtedly benefit from further refinement involving item analysis procedures. Finally, it should be noted that all the hypotheses of the present study were tested by semiprojective test data rather than by overt behavior in life situations. The findings are, therefore, limited to whatever extent these two classes of behavior do not directly covary. It is obvious that perfect covariation of test behavior and overt behavior is unlikely, especially where punishment frequently follows the responses in life situations. Nevertheless, the present hypotheses would be maintained—and are testable—in behavioral situations where factors likely to inhibit free responding are controlled or minimized.

Our orientation to this research was that, in predicting the behavior of individuals on the basis of presumed personality dimensions such as authoritarianism, no general conclusions should be drawn without taking into account the specific situational cues that may mediate this behavior. Within the limitations of this study, we have shown that punitiveness is not a general tendency that may be identified with authoritarianism, but rather that the punitive behavior of an individual in an interactional situation is dependent upon both his

authoritarianism and certain social status cues present in his stimulus environment.

SUMMARY AND CONCLUSIONS

Some of the interrelationships among authoritarianism, punitiveness, and status were studied by means of a semi-projective technique based on the *Rosenzweig Picture-Frustration Study*. It was found that when punitiveness is measured by this technique, high and low authoritarians do not differ with respect to extrapunitiveness independently of status. When social status is considered, however, high authoritarians, more often than low authoritarians, tend to respond with personal hostility to frustrators of low status and with indirect or displaced hostility to frustrators of high status. The data with respect to our hypothesis that low authoritarians respond to frustrators independently of their status are somewhat ambiguous but generally affirmative. We conclude that punitiveness is a function of both the personality of the individual and the particular stimulus environment that elicits the behavior.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. GUBA, E. G., & GETZELS, J. W. Frustration reaction and authoritarianism. *Amer. Psychologist*, 1956, 362. (Abstract)
3. LINDZEY, G. An experimental test of the validity of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 315–320.
4. LINDZEY, G. Differences between the high and low in prejudice and their implications for a theory of prejudice. *J. Pers.*, 1951, **19**, 16–40.
5. ROBERTS, A. H., & JOHNSON, D. M. The relationship of ethnocentrism to punitiveness as measured by the Rosenzweig Picture-Frustration Study. Paper read at Midwest. Psychol. Ass., Cleveland, 1952.
6. ROSENZWEIG, S. The picture association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, **14**, 3–23.
7. ROSENZWEIG, S., FLEMING, E., & CLARKE, H. Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, **24**, 165–208.
8. WILLIAMS, H. H. An investigation of extrapunitiveness as measured by the Rosenzweig Picture-Frustration Study and the California scale of the antidemocratic personality. Unpublished master's thesis, Univer. of Pittsburgh, 1949.

Received May 10, 1957.

CONFORMITY AND PROBLEM SOLVING[1]

CHARLES Y. NAKAMURA

University of California, Los Angeles

RECENT studies have expressed considerable interest in the relation between problem solving and various nonintellectual personality variables. Impetus in this direction has been provided, in part, by the fact that in problem-solving tasks individual differences are repeatedly found that are not accounted for by differences in general intelligence as measured by the conventional tests. In the search for factors that contribute to this variability, a number of experimenters have investigated nonintellectual variables (6, 12, 13). Although by no means conclusive, these studies present considerable evidence that such variables do contribute to individual differences in problem-solving performance.

This experiment was undertaken to investigate the possibility that conformity, a nonintellectual variable, may be shown to account for variation in problem-solving scores that is not accounted for by scores on a conventional intelligence test. In this study, conformity refers to behavior that is induced by subjecting individuals in a laboratory setting to conflict-producing social pressure. The experimental situation is a modification of that originated by Asch (2). The typical Asch situation is one of conflict for the naive subject who must either respond in accord with his own perception, thus disagreeing with the majority opinion, or respond in agreement with the group contrary to his own perception. An individual is said to conform if he resolves the conflict by agreeing with the contrived group consensus. This concept of conformity is to be distinguished from conformity viewed as normative or typical behavior, as in Floyd Allport's classical J-curve studies (1).

Although a survey of the literature revealed no direct investigation of problem solving and conformity, indirect but relevant information is found in the fact that similar personality traits correlate with both. Studies by Rokeach (15) and Brown (4) have demonstrated that rigidity in problem solving, measured by the Luchins water-jar test, is related to such variables as ethnocentrism and authoritarianism as measured by the California E and F scales respectively.[2] Conformity, too, has been shown to be significantly related to these same or similar variables in experiments by Crutchfield (7) and Hoffman (9). These studies provide support for the assumption that rigid problem solvers and high conformers have certain personality characteristics in common.

The present study was designed to test three hypotheses. The major hypothesis posits a negative correlation between tendency to conform and achievement in problem solving, when the influence of intelligence is statistically removed. The use of partial correlation permits a demonstration that conformity is associated with a portion of the variability in problem-solving performance that is not accounted for by differences in intelligence as measured by the conventional tests. It is generally accepted that general intelligence is a factor in problem-solving ability. Crutchfield (7) has reported a correlation of $-.51$ for a verbal test of intelligence and conformity.

The second hypothesis states that the negative partial correlation is higher for problems that require restructuring than for straightforward problems in which restructuring is unnecessary. This elaboration is plausible if

[1] This is a condensation of a report which was part of a general program of research on thinking in the Department of Psychology at Stanford. It was carried on under Project NR 150-149 and partially supported by Contract N6onr 25125 between Stanford University and the Office of Naval Research. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by or for the United States Government. The original report has been submitted to the committee on Graduate Study of Stanford University as a dissertation in partial fulfillment of the requirements for the Ph.D. in psychology. The writer is indebted to Donald W. Taylor for his support throughout the study. Thanks are due to Kenneth B. Little and Quinn McNemar for their suggestions and to Olga W. McNemar for editing the original manuscript. The writer is especially grateful to Richard S. Crutchfield for making available the research facilities of the Institute of Personality Assessment and Research, University of California, Berkeley.

[2] Since this study was designed and carried out, Levitt (11), in a recent review of the literature, has presented strong evidence that the validity of the water-jar test as a measure of rigidity is highly questionable.

one assumes that the water-jar Einstellung task—performance on which has indirectly been related to conforming behavior—has more in common with the restructuring type of problem than with the straightforward type, in that the two former require the modification of an initial approach in order to arrive at an adequate solution. This flexibility in approach requires additionally a critical evaluation of the information presented. It is assumed that these characteristics are also involved in the process of making independent judgments in the conformity situation.

The third hypothesis is that the sex differences found for certain types of problem solving are associated with sex differences in conformity. Several experiments (3, 8, 16) have demonstrated that men are superior to women in particular types of problem solving. Sex differences in conformity (7, 17) and in suggestibility (5) have also been demonstrated.

METHOD

Subjects

The *S*s were 64 men and 77 women students in the introductory psychology course at the University of California, Berkeley. The age range was 16 to 21 years for the women and 18 to 36 years for the men with respective median ages of 18 and 19.

Measures

Problem solving. The problem-solving measure consisted of 20 problems in booklet form. On the basis of an item analysis, the individual problems for this test were selected from those originally used by Sweeney (16) and Hunt and Taylor (10), so as to provide measures of achievement on two types of problems, restructuring and straightforward. Problems of the latter type were workable by direct methods. The former were problems in which for most *S*s the initial set or perception of the problem was inappropriate and a change in set or restructuring was required before solution could be attained. Ten straightforward and ten restructuring problems were arranged alternately and in order of difficulty. A working time of four minutes per problem was allowed with a ten-minute intermission after the first ten problems. Problems were scored either correct or incorrect, with a correct answer receiving one point credit. No partial credits were given.

Conformity.[3] The scale used consisted of 97 items mounted on individual slides. There were 52 critical items which were divisible into the following categories:

24 objective items; 4 each of 6 types
16 attitude items; 4 each of 4 types[4]
12 preference items; 4 each of 3 types.[5]

A response was scored as conforming if it deviated from the objectively correct response in the direction of the supposed responses of the other members of the group. A conforming response received one point and other responses none. The maximum possible score thus was 24, the number of items on the objective conformity scale.

The apparatus for measuring conformity involved five adjacent electrical panels on each of which were five rows of 11 signal lights and a row of 11 switches. The five rows of lights were lettered A, B, C, D, and E. The panels were separated by side wings which prevented each *S* from seeing any panel except the one in his own booth. The five panels were wired separately to the *E*'s control panel, located behind the *S*s, in such a way as to permit the *E* to send identical information to all five panels simultaneously. The *S*s' responses were received on a section of the *E*'s control panel by way of five rows of 11 signal lights, each row corresponding to a booth.

General Intelligence.[6] Part II, Analogies of Form B of the Concept Mastery Test[7] was selected as the measure of general intelligence. There are 75 multiple-choice analogy items; the score is the number right minus one-half the number wrong. The test was administered during one of the regularly scheduled lecture periods.

Procedure

All the students in the class of approximately 650 were informed that they were required to participate as *S*s in two sessions of psychological experiments. For this experiment, *S*s registered for a two-hour first session in groups no larger than 25 persons. The groups were composed without regard to sex, and the *S*s were seated around tables partitioned so that they could not

[3] Except for slight modification of the content and the order of presentation of the items, the method, materials and apparatus for measuring conformity were those designed and described by Crutchfield (7). They were used for this experiment with his cooperation.

[4] No further description of the attitude items will be presented since scores on them were not included in the final analysis of the data for reasons to be given below. A detailed description of the items, procedure of administration and system of scoring can be found in the original report (14). Copies of this report and also of that by Carey (6) and by Sweeney (16) are available in each of more than twenty major university libraries.

[5] These items are not included in the analysis of this experiment since the "normally" expected responses to them had not been established. The items were used by Dr. Crutchfield who employed the same subjects for another experiment.

[6] There was also available a score for each *S* on Form A of the College Vocabulary Test. This test was developed by H. G. Gough, Institute of Personality Assessment and Research, University of California, Berkeley. The relation of these scores to some of the experimental variables was included in the original report (14) for comparative purposes. However, the results will not be presented here since in the major analysis of the data, only the Analogies test scores were treated as the measure of general intelligence. The correlation between the two tests was .77 for the men and .68 for the women.

[7] The Concept Mastery Test was used with the kind permission of Lewis M. Terman, Stanford University.

see the papers of other *Ss*. At the end of this first meeting, devoted to the problem-solving task, *Ss* were asked to sign up for a second session, approximately an hour and a half in length, in groups of five men or five women.

At the second session, the *Ss* were seated at their respective panels in individual booths. The procedure was explained to them in some detail. About midway of the instructions, the *Ss* were shown five cards each lettered with an A, B, C, D, or E. They were told that they were to draw cards to determine the order in which the five *Ss* were to judge the stimuli as they were projected onto the screen.

The card-drawing took place immediately preceding the presentation of the first item to be judged. Before the *Ss* drew for letter designations, the cards were covertly switched, so that every card carried the letter E and each *S* was thus led to believe he was to respond last in the E position. The responses he saw appearing in the rows on his panel supposedly corresponding to *Ss* A, B, C, and D were actually predetermined, simulated responses transmitted by the experimenter from the control panel. The five *Ss* thus experienced identical stimulus conditions and gave their responses at the same time. To illustrate with one of the objective items: A slide was projected on the wall in front of the five *Ss*, containing a circle and a star and instructions to judge which figure had the greater area. Each *S* saw the other four simulated responses—all indicating the star as larger in area—before giving his own response. These simulated judgments were contrary to each *S*'s perception, since the circle was in fact about one-third greater in area than the star. On all critical slides, pressure to conform was induced by presenting a majority consensus that was inconsistent with the *Ss*' perceptions.

At the end of the second session, the experimenter discussed the study with the *Ss*, explaining the nature of the deception.

RESULTS AND DISCUSSION

The reliability coefficients, means, and standard deviations for the problem-solving and conformity measures for men and women are presented in Table 1. Considering the nature of the problem-solving test, the reliability coefficients of .70 and .76 for total scores were about as high as could be expected, although higher reliability for the restructuring and straightforward subtests would have been desirable. The reliability values for the objective conformity scores, .87 and .80, were only slightly lower than the .90 reported by Crutchfield (7) for total scores including not only objective, but also attitude and preference items. However, the reliabilities of .29 and .13 obtained for men and women respectively on the attitude conformity items were so low that no further analysis of data based on these items appeared justified. On the basis of unpublished data obtained earlier by Crutchfield, it had been expected that such attitude items would yield a considerably more reliable scale. Just why this expectation was not confirmed is not clear. In any case, all subsequent analyses were limited to conformity as measured by the objective items.

The correlation between the restructuring and the straightforward problems was .59 for men and .58 for women.

Within Groups Analysis

Table 2 presents the correlations among the measures of problem solving, conformity, and general intelligence.

TABLE 1
RELIABILITY OF THE PROBLEM SOLVING AND CONFORMITY MEASURES
(64 men and 77 women)

Measure	Group	r	M	SD
Total problems	M	.70	8.83	3.78
	W	.76	6.00	3.45
Restructuring problems	M	.54	4.33	2.16
	W	.60	2.79	1.95
Straightforward problems	M	.56	4.50	2.08
	W	.52	3.21	1.95
Attitude conformity	M	.29	8.63	3.11
	W	.13	9.75	3.17
Objective conformity	M	.87	8.44	5.09
	W	.80	10.39	4.70

Note.—Reliability was evaluated by the odd-even method for the total problems and the objective conformity scores. The first half versus the second half method was used for the attitude conformity scores. The r values were corrected for length by the Brown-Spearman formula. Reliabilities for the restructuring and the straightforward problems were computed by the Kuder-Richardson formula.

TABLE 2
CORRELATIONS BETWEEN MEASURES OF PROBLEM SOLVING, CONFORMITY, AND GENERAL INTELLIGENCE
(64 men and 77 women)

Measure	Objective Conformity		Analogies	
	M	W	M	W
Total problems	−.31**	−.18	.35**	.26**
Restructuring problems	−.38**	−.14	.35**	.22*
Straightforward problems	−.19	−.18	.27*	.25*
Analogies	−.31**	−.08		

* Significant at the .05 level.

** Significant at the .01 level.

Note.—The p values are for a one-tailed test since the directions of the correlations were predicted. A one-tailed test was used in all of the subsequent tables, except where otherwise indicated.

For men, conformity was found to be significantly negatively correlated both with total problems and with restructuring problems, but not with straightforward problems. After intelligence was partialled out, the correlations for conformity with restructuring problems (−.28) and with total problems (−.22) were significant at the .01 and the .04 levels. The z' transformation was used to test the significance of these partial correlations. These results support the first hypothesis: there is a negative correlation between tendency to conform and achievement in problem solving when the influence of intelligence is statistically removed. This finding indicates that conformity accounts for variation in problem-solving scores in addition to that which can be attributed to differences in intelligence.

For women, the zero-order correlations between problem solving and conformity were negative; they were, however, not significantly different from zero and hence cannot be regarded as supporting the research hypothesis. On the other hand, these correlations do not provide grounds for rejecting the hypothesis in that they did not differ significantly from those for the men.

The results for the men, for whom the restructuring problems were significantly negatively related to conformity whereas the straightforward problems were not, are compatible with the second hypothesis, that the negative partial correlation is greater for the restructuring problems than for the straightforward problems. However, the difference between the partial correlations was not significant ($p = .09$). Negative results were also obtained for the women.

An inspection of the scores for men revealed that the older men tended to score higher on problem solving and analogies but lower on conformity than the younger men in the sample. For this reason, separate analyses were made for the group of fifty men 21 years of age or younger and for the group of 14 men ranging in age from 22 to 36. These results are shown in Table 3. For the fifty younger men, only the correlation between restructuring problems and conformity was significant. On the whole, the magnitude of the correlations for the younger men differed little from that for the women.

For the 14 men aged 22 and over, there were high negative correlations between problem solving and conformity. After the effect of intelligence was partialled out, the correlation was −.61 ($p = .02$) for total problem solving versus conformity and − .76 ($p = .005$) for restructuring problems versus conformity. With the influence of intelligence and age controlled simultaneously, the respective partial correlations were −.59 ($p = .03$) and −.77 ($p = .005$). The z' technique was used to test the significance of these partial correlations.

The results in Table 3 show also that the difference between the correlations of restructuring and straightforward problems with conformity was in the hypothesized direction. However, the difference between the partial correlations was again not significant. For both the young and the older groups p was larger than .10. Thus, the analysis of the scores for the older men clearly supports the first hypothesis of the study but not the second hypothesis.

Between Groups Analysis

The third hypothesis was that sex differences in certain types of problem solving are associated with sex differences in conformity. It was reasoned that cultural factors that impose differential experiences, training, pressures, and expectations would result in differences in motivations and skills for the two sexes. It was expected that these differences would be manifested in performance on certain problem-solving tasks and under con-

TABLE 3

CORRELATIONS BETWEEN MEASURES OF PROBLEM SOLVING, CONFORMITY, AND GENERAL INTELLIGENCE

(50 young men [Y] and 14 older men [O])

Measure	Objective Conformity		Analogies	
	Y	O	Y	O
Total problems	−.20	−.70**	.25*	.65**
Restructuring problems	−.25*	−.79**	.21	.70**
Straightforward problems	−.11	−.59**	.25	.38
Analogies	−.22	−.44		

* Significant at the .05 level.

** Significant at the .01 level.

Note.—The z transformation was used to test the significance of the correlations for the group of 14 older men.

formity conditions similar to those of this experiment.

For this hypothesis to be valid first requires that men be superior to women in certain types of problem solving and at the same time show less tendency to conform. Table 4 shows that these prerequisite conditions were clearly fulfilled. The sex differences for all three problem-solving scores were highly significant. However, the sex difference on restructuring problems was not significantly greater than that on the straightforward problems. The sex difference on the conformity measure was significant as expected. Surprisingly, a significant sex difference was also obtained on general intelligence. Because of this unexpected difference, it was necessary to adjust for both intelligence and conformity in carrying out the analysis of covariance designed to test the hypothesis. The results of the analyses, carried out separately for the restructuring and the straightforward problems, are presented in Table 5. Although there was some reduction in the differences, both F ratios were still significant at the .01 level, indicating that not all of the sex differences in problem solving was accounted for by sex differences in conformity and intelligence as measured.

A similar analysis of covariance was not carried out for the sex difference on total problems. Since significant differences remained after adjustments on the restructuring and the straightforward problems separately, it was clear that the difference would be significant for total problems should adjustments be made for conformity and intelligence.

The question may be raised as to the relevance of the results of the separate covariance analyses of the restructuring and straightforward problems to the earlier findings of Sweeney (16). In a series of experiments Sweeney found that for groups matched with respect to intelligence, special abilities, or specific knowledge, men were superior to women on restructuring types of problems but not on straightforward problems. In the present experiment men proved to be superior to women in both types of problems after the covariance adjustments were made for conformity and intelligence. Though these results do not confirm Sweeney's findings, it cannot be said that they are contradictory since the groups in the present study were not equated on relevant special abilities or on related specific knowledge.

TABLE 4
SEX DIFFERENCES
(64 men and 77 women)

Measure	Group	M	SD	D	p
Total problems	M	8.83	3.78		
	W	6.00	3.45	2.83	$<.001$
Restructuring problems	M	4.33	2.16		
	W	2.79	1.95	1.54[a]	$<.001$
Straightforward problems	M	4.50	2.08		
	W	3.21	1.95	1.29[a]	$<.001$
Conformity	M	8.44	5.09		
	W	10.39	4.70	1.95	.01
Analogies	M	37.19	12.17		
	W	29.58	10.68	7.61	$<.001$[b]

[a] The difference between mean differences in restructuring problems and straightforward problems is .25; p is not significant.
[b] The p value is for a two-tailed test since the difference was not predicted.

TABLE 5
ANALYSES OF COVARIANCE: SEX DIFFERENCES ON RESTRUCTURING PROBLEMS AND ON STRAIGHTFORWARD PROBLEMS, SEPARATELY, WITH ADJUSTMENTS FOR CONFORMITY AND ANALOGIES SCORES

Test	Source	Before Adjustment		After Adjustment			
		df	Mean Square	df	Mean Square	F	p
Restructuring problems	Total	140	4.79	138	4.00		
	Within	139	4.24	137	3.79		
	Between	1	32.00	1	32.44	8.56	.01
Straightforward problems	Total	140	4.41	138	3.99		
	Within	139	4.09	137	3.79		
	Between	1	58.00	1	30.71	8.10	.01

A further analysis of sex differences on problem solving was made to determine whether or not sex differences would remain if the older men were excluded. The differences between the means for 77 women and 50 younger men were computed for the problem solving, conformity, and analogies scores. The differences on all three measures of problem solving remained significant at the .001 level. The scores of the older men were thus not entirely responsible for the obtained sex differences. For conformity, the difference was reduced in level of significance from .009 to .07. The difference on analogies was still significant at the .003 level. An analysis of covariance with adjustments for conformity and intelli-

gence was not carried out for this comparison as it was for the difference in problem solving for the total men and the women. Since the difference between the means for the younger men and the women in conformity was very small and since the correlations between conformity, problem solving, and intelligence scores for the younger men were less than the corresponding correlations for the total men, covariance adjustments would not be expected to reduce the sex difference in problem solving to insignificance.

SUMMARY

This study was undertaken to investigate the possibility that conformity, a nonintellectual variable, contributes to the variability in achievement in problem solving in addition to that variation accounted for by intelligence as measured by the conventional tests. Three hypotheses were tested:

1. There is a negative correlation between tendency to conform and achievement in problem solving when the influence of intelligence is statistically removed.

2. The negative partial correlation is higher for problems that require restructuring than for straightforward problems.

3. The sex differences in certain types of problem solving are associated with sex differences in conformity.

The *S*s, 77 women and 64 men, were given tests of problem solving, conformity, and intelligence. The problem-solving test consisted of 10 restructuring and 10 straightforward problems. Part II, analogies, of Form B of the Concept Mastery Test was used as the measure of intelligence. Two measures of conformity were used, one involving judgments concerning objective stimuli and the other judgments concerning attitudes. Since the reliability of the latter proved to be extremely low, all subsequent analyses were limited to conformity as measured by the objective items.

The results supported the first hypothesis but failed to confirm the second and third hypotheses.

REFERENCES

1. ALLPORT, F. H. The J-curve hypothesis of conforming behavior. *J. soc. Psychol.*, 1934, **5**, 141–183.
2. ASCH, S. E. Studies of independence and conformity: 1. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, No. 9 (Whole No. 416).
3. BILLINGS, M. L. Problem solving in different fields of endeavor. *Amer. J. Psychol.*, 1934, **46**, 259–272.
4. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
5. BROWN, W. Individual and sex differences in suggestibility. *Univer. California Publ. Psychol.*, 1916, **2**, 291–430.
6. CAREY, G. L. *Reduction of sex differences in problem solving by improvement of attitude through group discussion.* Stanford: Dept. Psychol., Stanford Univer., 1955 (Tech. Rep. No. 9, Contract N6onr 25125).
7. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
8. GUETZKOW, H. An analysis of the operation of set in problem solving behavior. *J. gen. Psychol.* 1951, **45**, 219–244.
9. HOFFMAN, M. L. Some psychodynamic factors in compulsive conformity. *J. abnorm. soc. Psychol.*, 1953, **48**, 383–393.
10. HUNT, E. B., & TAYLOR, D. W. Personal communication. January, 1955.
11. LEVITT, E. E. The water-jar Einstellung test as a measure of rigidity. *Psychol. Bull.*, 1956, **53**, 347–370.
12. LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
13. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
14. NAKAMURA, C. Y. *The relation between conformity and problem solving.* Stanford: Dept. Psychol., Stanford Univer., 1955 (Tech. Rept. No. 11, Contract N6onr 25125).
15. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
16. SWEENEY, E. J. *Sex differences in problem solving.* Stanford: Dept. Psychol., Stanford Univer., 1953 (Tech. Rep. No. 1, Contract N6onr 25125).
17. WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

Received May 13, 1957.

A RESTATEMENT AND TEST OF SCHLOSBERG'S THEORY OF EMOTION WITH TWO KINDS OF SUBJECTS FROM GREECE[1]

HARRY C. TRIANDIS

Cornell University

AND WILLIAM W. LAMBERT

Center for Advanced Study in the Behavioral Sciences

MANY propositions in the literature of psychology are stated in such a general form that replication with human subjects from very diverse cultural backgrounds is directly relevant. In the present state of our understanding, such replication is probably required whenever a proposition reflects experimental operations dependent upon the more subtle uses of perceptual or language responses. If the proposition is upheld by the test, then the possibility that the relation depends on rare cultural or social factors is rendered less likely.

This study represents an attempt to test Schlosberg's dimensional theory of emotion (15) and to compare the relationships found using American college *Ss* with those found with Greek college students from Athens and *Ss* from a small Greek village. Besides the test of the main hypotheses relating to similarities in the data, certain exploratory hypotheses concerning differences are also discussed.

Schlosberg (14) has attempted to reduce the dimensions required to order judgments of pictures that presumably reflect "emotional states" in the person photographed. Woodworth (17) had *Ss* place pictures into categories named: 1. love, happiness, mirth; 2. surprise; 3. fear, suffering; 4. anger, determination; 5. disgust; 6. contempt. Schlosberg has shown that it is reasonable to view this continuum as a circle rather than a straight line. Further, he has shown that there is a high correlation between the way *Ss* place the pictures into Woodworth's six categories and the way they place the same pictures according to the two bipolar nine-step dimensions of: 1. pleasant-unpleasant; 2. attention-rejection. These dimensions are fewer in number than those of Woodworth and they possibly involve somewhat less complex judgments. However, much of their face validity and their relation to the extensive everyday language realm of "emotional states" depends largely on the even more subtle and complex discriminal and symbolic processes involved in the use of the Woodworth categories. More recently Schlosberg (15) has extended his theory to include a further dimension, tension-sleep. The appropriate model thus becomes an inverted ellipsoidal cone, with "sleep" anchoring the base point of the cone.

Although this latest version of the theory is still primarily concerned with the *perception* of emotional expression, the physiological indices used by Schlosberg for the measurement of the tension-sleep dimension suggest that the theory is a first step towards a general theory of emotion. One can think here of Hebb's (2), and of Lindsley's (4) discussion of the differential firings in the reticular formation as the physiological changes underlying the *sleep-tension* dimension, of Olds's (9) firing of the "reward system in the brain" (the septal area facilitation and inhibition) which suggests the basis for the *pleasant-unpleasant* dimension, and of the action of the sympathetic and parasympathetic nervous systems, with their dilation and constriction of the pupil of the eye and the bronchi etc., which suggest a basis for the *attention-rejection* dimension. In short, although the present data involves judgments of facial expressions, it may be possible to generalize the *dimensions* abstracted from these data to subsume a good portion of the variance of the actual emotions people feel. Certainly when one judges the emotion shown in a face, the particular judgment may not correspond either to one's own emotion at the time or to that of the person who is moving his face. The dimensions of the judgment, however, may turn out to be adequate to describe both one's own and

[1] This study was assisted in part by a grant from the U. S. Public Health Service to the second author and by the general research funds of the Department of Psychology, Cornell University. Writing and analysis was in part facilitated by the Center for Advanced Study in the Behavioral Sciences. The second author is on leave from Cornell University. The authors are appreciative of suggestions given by A. T. Ryan and J. E. Hochberg.

the other's emotional states. Proof of such an hypothesis, however, will depend on both the development of more data on judgments as well as more independent physiological measurements, and, most important of all, development of experiments relating both of these.

Osgood (11) has recently reported the results of a factor analysis of *Ss*' responses to forty posed facial "expressions," using the semantic differential as a measurement device. He suggests that the factors he named after applying Thurstone procedures to the picture judgments—1. pleasantness; 2. control; 3. expressiveness;—"look like" the major dimensions of his "semantic space"—1. evaluation; 2. potency; 3. activity. This appears to be a reasonable suggestion, and it only remains for us to suggest that Schlosberg's three dimensions appear to be either a special case or a biologically more generic case of the dimensions later named by Osgood.[2]

Keeping in mind the problems of factor naming, or even of dimension naming, one may suggest that Schlosberg's "pleasant-unpleasant" sounds like Osgood's "pleasantness" and "evaluation," and "tension-sleep" sounds like "activity" and "expressiveness," whereas "attention-rejection" sounds like "potency" (in the sense of potency of the cause of emotion) and more clearly, "control."

If such convergence turns out to be the case it would be interesting. A rather nice question would then arise as to which set of operations might be most profitably viewed as more generic. The facial expression operations have one advantage over the more complicated "semantic" ones in that they keep open the possibilities of a direct psychophysical approach to the further clarification of the matter. Although symbol capacity and usage is certainly of vital importance in human behavior, the basic and most general dimensions of such capacity and usage may be presaged in some psychophysical predispositions common with some other animals and with all peoples. Teasing out the parameters involved in the stimuli of "emotional expressions" may be the most direct attack on the bases of the more elaborated and secondary semantic mediating processes.

[2] The editor of this Journal has suggested that our formulation is quite parallel to Wundt's tridimensional theory of feeling.

Since there is no basis a priori for assuming that cues for discriminating "facial expressions" are more universal or generic than the apparently more complex semantic dimensions, both formulations require tests in very different cultural milieu before any general formulation of either theory be taken very seriously. Kumata (3) and Triandis and Osgood (16) have attempted cross-cultural tests of the semantic operations and their attendant hypotheses. The present paper reports such an attempt for the facial "expressions" operations. As a result of such attempts we may be able to take seriously the partial but general pattern that seems to be emerging from all of these studies.

This possibility is all the more interesting when the similar convergence in the independent work of Morris (7) on dimensions of "values" and their setting in a theory of behavior (8) is considered. Morris and Jones (6) have factor analyzed the relations between the ratings of thirteen "ways of life." They report five Thurstone factors, three of which can be viewed as relevant here. Their "receptivity and sympathetic concern" factor "looks like" a meeting of the "sleep" or "passive" end of the "sleep-tension" dimension and of the "acceptance" end of the "acceptance-rejection" dimension. Their "enjoyment and progress through action" factor could be a meeting of the "active" and "approach" ends of the Schlosberg dimensions, and finally, their "withdrawal and self-sufficiency" factor could be a meeting of the "withdrawal" and "rejection" ends of our continua.[3] This suggestion, of course, does no more than outline the *possibility* of a convergence of the Schlosberg dimensions and of the Osgood "meaning" dimensions with basic dimensions of "valuing" behavior.

METHOD

Subjects

The American data have been kindly provided by T. Engen, N. Levy, and H. Schlosberg, based on ninety Brown University students. Two groups were tested in Greece during the summer of 1956. Fifteen *Ss* were studied in the village of Sfakera on the northeast coast of the island of Corfu, and fifteen college students from Athens. In both groups there were almost an equal number of *Ss* of each sex. Other characteristics of the two groups are given in Table 1, which makes clear that there are large differences between the groups in

[3] These suggestions were made by Fred Attneave in personal communication.

TABLE 1
CHARACTERISTICS OF SAMPLES

Statistic	Age		Education (years)		Movie Attendance (lifetime)	
	V	A	V	A	V	A
Mean	18.6	21.1	7.1	14	19	1000
Median	19	18	6.0	13	12	900
Range	15–24	17–33	6–11	11–18	0–50	100–3000

Note.—V stands for village and A for Athens.

educational level and in their attendance at movies—largely the same movies American students would attend. The data on movie attendance in the table represent estimates by the *S*s themselves.

Procedure

The first problem was to ensure equivalence of terms in Greek and English. This was approximated by having a group of bilinguals translate the terms to be used from Woodworth's and Schlosberg's English into Greek, and another group of bilinguals translate the Greek back into English. Only terms which consistently met a criterion of complete agreement both ways were accepted for the Greek version.

In the Schlosberg procedure, anchoring pictures were presented while instructions were given, and then taken away before an *S* began rating. Pictures numbered 8 and 32 of the Ruckmick (13) series were used for the pleasant-unpleasant dimension, Numbers 7 and 13 for the attention-rejection dimension and 26 and 4 for the tension-sleep dimension.

The pictures judged by the *S*s were the new set of pictures of Marjorie Lightfoot, kindly provided by Schlosberg, which are reproduced in Morgan's workbook (5, p. 32–33).

The nine-point scales and the Woodworth six categories were drawn on cardboard for presentation to the *S*s.

The villagers were instructed as follows:

> This is an important study which I am making, but as far as you are concerned it is just a game. All you have to do is to play the game *correctly*. The rules of the game are as follows: you are to place each of these forty-eight photographs in one of these places (show on cardboard). First you will place those that represent a pleasant emotion on *this* part of the cardboard (indicate). Then you will place those that represent an unpleasant emotion on *that* part of the board (indicate). Then you will spread them out, so that the more pleasant ones are there, and they gradually get less and less pleasant, up to here (indicate) and more and more unpleasant from here (indicate) to there.

Similar instructions were given for the other dimensions. In order to make sure that the idea "gradualness" was communicated a pencil was tapped on a table at various intensities. In all cases in the village, the *S* was not out of sight of his peer group, but was sufficiently far from his friends so that they would not influence his ratings. The villagers were very eager to "play the game" and they all said that they "enjoyed playing it." Token gifts were given to all participants and refreshments were served in all instances. This kind of non-business-like atmosphere is thought to be essential for psychological research in Greek villages.

In the case of the city group the instructions were the same, but the task was presented as an experiment, and a test of their ability to recognize emotions.

RESULTS AND DISCUSSION

Similarity Between Ratings of Greek and American Subjects

The first hypothesis was that Greek *S*s rate emotional "expression" in much the same way as American *S*s. A test of this hypothesis requires that there be a significant and fairly high correlation between the way the three groups place the pictures on the three Schlosberg dimensions. Such a test is shown in Table 2. Rank-order correlation coefficients were computed on the ranks of the forty-eight pictures. The lowest correlation coefficient in this table is .66 and is significant at the .0001 level. Hence, there is little doubt that Greek *S*s, even when they come from very different populations, rate emotional expressions in the same way as American college *S*s.

A corollary of this hypothesis is that if the P-U, A-R, and S-T are assumed to be orthogonal, and the position of the photographs plotted accordingly, the *emotion solid* obtained will be similar for the Brown University and Greek data. This hypothesis can be verified by visual inspection of Fig. 1. These solids are plotted by the method of successive intervals, a scaling procedure derived from Thurstone's law of comparative judgment (see Edwards [1, pp. 120–135]).

Briefly, we obtain all available pictures of a certain level of tension, say between 7 and 8 (on a nine-point scale), and note how they have been rated on the A-R, and P-U scales. We then treat each picture just as Edwards

TABLE 2
RANK-ORDER CORRELATION COEFFICIENTS

Groups	Dimension		
	P-U	A-R	S-T
Brown vs. Athens	.91	.82	.76
Brown vs. Village	.82	.88	.67
Athens vs. Village	.86	.83	.66

Note.—Based on ranks of forty-eight pictures showing relations among three groups studied.

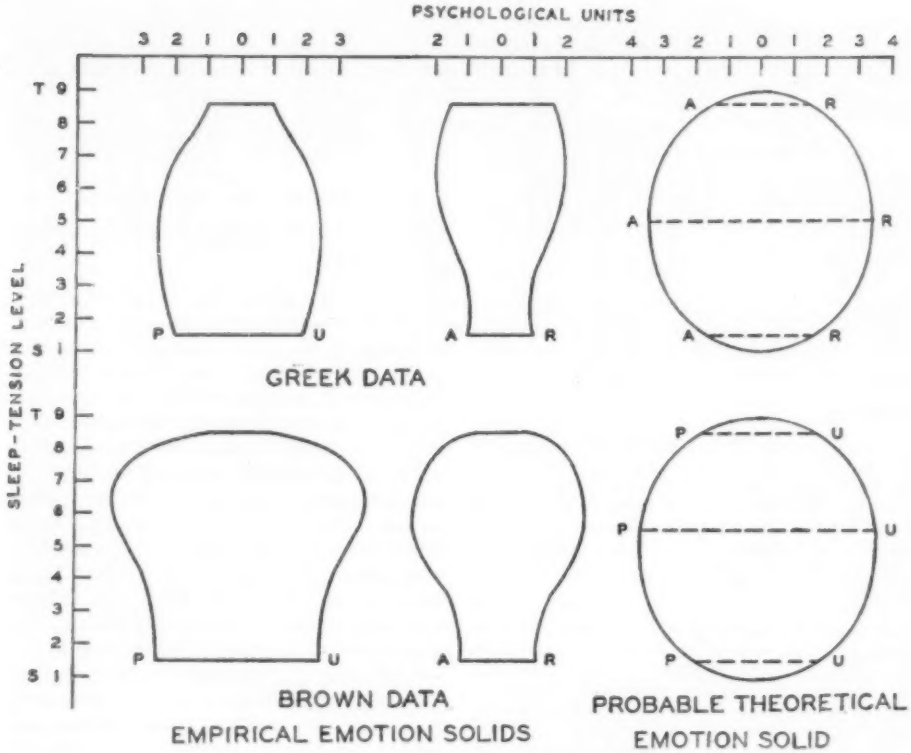

FIG. 1. EMPIRICAL AND THEORETICAL EMOTION SOLIDS

treats a *statement* for attitude scale construction: the frequency of judgments of this picture at a certain point of a psychological continuum is noted, this is converted to a cumulative distribution, and to a cumulative probability score. The latter is transformed to a normal deviate (z score) by referring to the normal probability distribution. The difference between adjacent intervals can thus be obtained. Repetition of this procedure for several pictures and averaging gives an estimate of the interval width for each one of the scale graduations of the P-U and A-R scale. If we add all the intervals, we get the length of the P-U and A-R axes. Since such axis-lengths can be computed at each level of tension we obtain several axis-lengths and use them to plot our solid.

We must not place too much confidence on the actual distances on these solids, since they are based on *only one* picture at the tension level between 1 and 2 and only 4 or 5 pictures for tension levels 3 to 4, 5 to 6, 7 to 8, and 8 to 9. These are small samples indeed of what are probably rather large populations. Furthermore, the American solid is based on more than 90 *S*s, the Greek on only 30 *S*s. Differences in the general width of the irregular ellipsoidal cylinders obtained may be due to better discriminations made by the greater number of *S*s for the American data. Nevertheless, we can cautiously conclude that the emotion solid is probably closer to an irregular ellipsoidal cylinder rather than an inverted ellipsoidal cone as was postulated by Schlosberg.[4] Schlosberg's guess on the shape of the solid was based

[4] The assertion that fewer discriminations can be made at the ends of the tension continuum than in the middle depends upon acceptance of the following assumptions: (*a*) a normal distribution of judgments for each picture (this is clearly not always true); (*b*) perfect test-retest reliability; (*c*) rating one picture does not influence the rating of another; (*d*) the distance between the poles remains constant throughout the series of judgments; (*e*) the poles mean the same thing to all *S*s.

on theoretical rather than empirical considerations. Our solids are completely empirical.

If a more smooth looking solid is desired, our best guess is to assume that the regular ellipsoid of Fig. 1 will be a reasonable approximation. Much more work is needed before the exact shape of the emotion solid can be determined. A serious attempt will have to be made to construct pictures which push at the outside contours of the solid, and more work must be done on scaling the sleep-tension levels themselves. Though the present method is laborious, it yields results that seem reasonably consistent even with very rough data.

Reliability of Judgment: Schlosberg Dimensions vs. Woodworth Categories

The second hypothesis was that Greek *Ss* will place pictures on the Schlosberg dimensions (pleasantness-unpleasantness, attention-rejection, and tension-sleep) more reliably than when they use the Woodworth categories. This hypothesis can only be given a rough test because of the difference in procedure in presenting the scales in the Schlosberg method (two polarities at a time) and that in the use of the Woodworth categories (six categories at once). Results are fairly clear from a careful "visual" check of the data. Two photographs in the case of the village group and one photograph in the Athens group were rated in what seems to amount to a random fashion on at least one dimension, when the Schlosberg procedure is used. In the use of the Woodworth categories, however, 24 of the pictures in the village and 11 in Athens, were allocated in what amounts to a random fashion. These findings suggest that the Schlosberg dimensions permit more reliable ratings of the photographs than do the Woodworth categories.

Predictability of Woodworth Categories from Schlosberg Dimensions

The third hypothesis was that the position in the Woodworth categories can be predicted from its rating on the pleasant-unpleasant and attention-rejection dimensions. We expect this prediction to be better in the case of the Athens group than in that of the village group since the Athens group is closer to the cultural inductive base—the United States—largely because of repeated attendance at American movies. To test this hypothesis the positions of the pictures as determined by the P-U and A-R dimensions were computed and plotted. The surface of the graph was divided into six equal sections, following Schlosberg's procedure. To permit a test of significance, a table such as sketched in Table 3 was prepared. The circularity of the Woodworth categories was assumed, since Schlosberg has shown that this is a reasonable assumption (14). According to Schlosberg, there should be a relationship between the placement in the six sections of the graph and the placement on the six qualitative categories of Woodworth. The "predicted category" is, therefore, the section of the graph in which a picture was placed, and the "mean actual category" is the mean placement in the Woodworth categories. If the actual category differed from that predicted by one or more categories, an error was counted. If it differed by 0.9 or less of a category, no error was counted. By this criterion, if the predictions were done by chance alone, two-thirds of the photographs would produce errors. Actually, fewer errors were produced. A chi-square test of goodness of fit between the predicted and the actual number of errors showed fewer errors with $p < .001$ for the Athens group, $p < .05$ for the village group. We conclude that the Greeks followed Schlosberg's predictions in showing a relationship between their placement of a picture on the P-U and A-R dimensions and in the Woodworth categories. These results provide some support for the third hypothesis.

TABLE 3
ERROR IN PREDICTION OF WOODWORTH CATEGORY PLACEMENT

Picture	Predicted Category	Mean Actual Category	Error
1	1	1.3	0
2	6	4.9	1
3	4	3.1	0
etc.	etc.	etc.	etc.
48	5	6	1

Note.—This is only a partial table. It is used in order to arrive at criterion of error of prediction for placement of Woodworth categories from P-U and A-R placement.

Relative Convergence of Groups with One Another

The fourth hypothesis dealt more specifically with the influence of stereotypy. Since the Athens and Brown *Ss* see many of the same

movies they ought to converge more with one another in their judgments and ratings of facial expressions than with the village group. A chi-square test was used again. For each picture and each dimension the differences in the mean ratings of the three groups were computed. If the Brown mean for a picture on a given Schlosberg dimension was 2.2, the Athens mean 1.9 and the village mean 2.9, the village *Ss* were classified as further from the other two since $2.9 - 2.2 > 2.2 - 1.9$, and the instance scored as one case of convergence between Brown and Athens. If the prediction of Hypothesis 4 is incorrect, there should be as many cases of Brown and Athens converging (16 cases) as there are cases of convergence between the other two possible pairings. Actually, this was not the case (see Table 4). The chi square for each dimension is significant at the .05 level. The overall chi square for the table is 11.4 or $p < .005$.

Thus we conclude that there is some factor bringing the Athens and Brown group closer together. A direct test of the relationship between movie attendance and closeness of judgments to the Brown norms was performed and was in the expected direction but did not quite reach significance.

Effects of Cultural Differences

The fifth hypothesis reflected a general expectation of some consistent and significant differences in judgments of the pictures associated with differences in customs. While the present data cannot exclude the possibility of chance association, such differences suggest interesting hypotheses that could be tested using these procedures in conjunction with sample surveys or anthropological methods.

Role of verbal aggression. It is the custom in the Greek village studied to engage in verbal aggression as a means of entertainment—a kind of aggressive play that may function to express anger but also provide practice in control of aggression. Loud and angry debates are a favorite pastime in Sfakera. It is interesting therefore to note that there are significant differences in judging the picture described by Schlosberg as "Intense anger in argument" (No. 22). Compared to either the Athens or Brown groups, the villagers see the picture as more pleasant and less tense (at the .05 and .01 levels respectively). The pictures included six other attempts to portray anger, and in all but one of these the villagers tended to judge them as more pleasant than did the Brown group. The only reversal with the Brown group was with "mild anger" states that might not reflect the dramatic quality enjoyed in their anger by these people. The differences in tension are less marked, but in the three cases where there are differences they reflect the same direction as found in picture No. 22. Differences between the Athens group and the villagers are not maintained in the other pictures, suggesting that a cultural factor may be operating. The hypothesis emerges, then, that to the degree that anger is displayed under pleasurable circumstances pictures of that state are viewed as more pleasurable and less tense. Such a hypothesis seems applicable to individual differences as well as to group or cultural differences.

TABLE 4
RELATIVE CONVERGENCE OF GROUPS TO ONE ANOTHER

Groups	P-U	A-R	S-T	Total
Brown and Athens converging	24	20	23	67
Brown and Village converging	11	12	14	37
Athens and Village converging	13	16	11	40
Totals	48	48	48	144

Role of bodily exertion. It is reasonable to expect differences in the judgment of pictures depicting physical strain between rural people who do physical toil and college students whose main contact with real strain may be limited to sports and games. Such differences are reflected in judgments of picture No. 30, which the villagers judge as less pleasant and yet less tense than do either of the college groups. There appears to be some cultural factor in addition since the same difference holds significantly for the Athens group compared to the Brown group. Certainly the Greeks generally depend much less on mechanical equipment for getting work done. Adaptation to physical strain may make its exemplification seem less tense, but the ubiquity of hard physical toil and the attendant fatigue should lead to association with pain and unpleasantness. The present data therefore, may provide a rough index of the valuation placed on physical toil and the effort expended, and we hypothesize that the valuation varies nega-

tively with frequency of exhausting labor and positively with frequency of games of physical skill, whereas adaptation to such strain is reflected in the effort or tension dimension. Here again individual differences as well as cultural differences should be reflected.

Relations among dimensions. One additional trend that runs throughout the data on the attention-rejection and tension-sleep dimensions is worthy of mention even though its interpretation is more difficult. On the average, the Greeks (whether urban or village) give ratings more toward "attention" and "tension" for *unpleasant* pictures while American college students give ratings more toward "attention" and "tension" for *pleasant* pictures (see Tables 5 and 6). These findings are significant at the .001 level by chi-square test.

In explanation of this finding we might invoke and apply to emotional expression a stereotype according to which Americans are seen as "activist" and "instrumentally oriented." According to the diagram used by Schlosberg (14, p. 127) for representing the relation between his dimensions and the Woodworth categories, the Americans, by giving higher "attention" ratings than the Greeks to pleasant pictures, shift them from the more purely emotional categories of "love, mirth, happiness" toward the more instrumental "surprise." Conversely, by giving relatively less "attention" ranging to unpleasant pictures, the Americans shift the latter from the more purely emotional categories of "fear, suffering" toward the instrumental categories of "anger and determination." The Greeks may be more ready to accept an emotion without having any idea of what to do about it. There is a suggestion that Americans tend to give higher "tension" scores on the average than do the Greeks in the present study.

One rather jarring note emerges from consideration of Picture No. 51. Seven out of fifteen villagers perceived it as "extremely pleasurable" while another six scored it as "extremely unpleasant." Both Athens and Brown groups saw it overwhelmingly as unpleasant (scores 1 to 3). This finding actually follows from the shape of the emotion solid. When an axis is shorter at one level of tension than at another discrimination at that level should be more difficult and confusion more likely.[5] Such confusion was typical in the case of extreme judgments of pictures that are judged to be toward either high tension or sleep.

TABLE 5

NUMBER OF PICTURES RELATIVELY HIGH ON "ATTENTION"

	Americans	Greeks	Total
Pleasant Pictures	12	3	15
Unpleasant Pictures	6	21	27
Totals	18	24	42

TABLE 6

NUMBER OF PICTURES RELATIVELY HIGH ON "TENSION"

	Americans	Greeks	Total
Pleasant Pictures	17	0	17
Unpleasant Pictures	9	15	26
Totals	26	15	41

SUMMARY

In the present paper we have shown that the conclusions presented by Schlosberg (1952, 1954) for Brown University subjects are valid in Greece as well. Schlosberg's method can now be used with non-Western groups, to establish the differences between the non-Western and Western modes of perceiving emotional expressions.[6]

Secondly, we have shown that the emotion solid can be plotted from data collected by making use of Schlosberg's procedure, and the shape of the solid is *not* the same as that presented in Schlosberg's 1954 paper.

Thirdly, we have shown that the Schlosberg technique can be used to study certain differences between Western cultural groups.

[5] This interpretation may be too simple. There is a highly significant difference in the greater number of bimodal distributions of judgments at the two extreme ends of the sleep-tension continuum as compared to the middle ranges. This fact, together with the statement of some *S*s that extremely pleasant and extremely unpleasant expressions look the same leads to the possibility that stimuli are more ambiguous at extreme tension and low tension levels and *also* that *S*s tend *not* to use the middle categories of the scales at these levels. Serious validation of these findings might suggest a more complicated solid than the one hypothesized in Fig. 1.

[6] Such a study is being planned for a village in Taiwan, to be conducted by Arthur and Margery Wolf.

REFERENCES

1. EDWARDS, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957.
2. HEBB, D. O. Drives and C.N.S. *Psychol. Rev.*, 1955, **62**, 243–254.
3. KUMATA, H., & SCHRAMM, W. A pilot study of cross-cultural meaning, *Public Opin. Quart.*, 1956, **20**, 229–237.
4. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: John Wiley, 1951. Pp. 473–516.
5. MORGAN, C. T. *Student's workbook to accompany introduction to psychology*. New York. McGraw-Hill, 1956.
6. MORRIS, C., & JONES, L. V. Value scales and dimensions, *J. abnorm. soc. Psychol.*, 1955, **51**, 523–535.
7. MORRIS C. *Varieties of human value*. Chicago: Univer. of Chicago Press, 1956.
8. MORRIS C. Signs, language and behavior. Paper read at Center of Advanced Study in the Behavior Sciences, January, 1957.
9. OLDS, J. Physiological mechanisms of reward. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. of Nebraska Press, 1955.
10. OSGOOD, C. Behavior theory and the social sciences, *Behav. Sci.*, 1956, **1**, 167–185.
11. OSGOOD, C. E. Fidelity and reliability. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill: Free Press, 1955.
12. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana, Ill.: Univer. of Illinois Press, 1957.
13. RUCKMICK, C. A. A preliminary study of the emotions. *Psychol. Monogr.*, 1921, No. 136.
14. SCHLOSBERG, H. The description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1944, **44**, 229–237.
15. SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.
16. TRIANDIS, H. C., & OSGOOD, C. E. A comparative factorial analysis of semantic structures in monolingual Greek and American college students. *J. abnorm. soc. Psychol.*, in press.
17. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
18. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Holt, 1954.

Received May 20, 1957.

PSYCHOLOGICAL TEST AND ORGANISMIC CORRELATES OF INTERVIEW INTERACTION PATTERNS[1, 2]

RUTH G. MATARAZZO, JOSEPH D. MATARAZZO, GEORGE SASLOW, AND JEANNE S. PHILLIPS

University of Oregon Medical School

In several previous papers, we have shown that certain behavioral variables which characterize an individual's interview interaction pattern during a partially standardized interview can be objectively and reliably assessed by the Interaction Chronograph method (7, 12, 13). This method of assessing personality through interview interaction patterns is based on the assumption that *how* a person interacts with an interviewer in regard to certain temporal interaction measures specified, for example, in terms of frequencies and durations of action, silence, and other variables, may be both characteristic and meaningful. That is, how often a person speaks, how long are his utterances, how long is his reaction time before responding to the interviewer, how often he takes the initiative when the interviewer is silent, etc., may reflect significant and predictive individual differences.

As described by Chapple (3) and by the writers (8, 12), the interviewer's behavior (but not the content of his communications) is standardized during the total interview, which consists of five subperiods. During Subperiods 1, 3, and 5, the interviewer's behavior consists of nondirective, free, give-and-take interviewing. Periods 2 and 4 are stress phases of the interview, which require the interviewer to remain silent following the patient's last utterance (Period 2) and to interrupt the patient while he is talking (Period 4) as a "dominance-submission" situation. The variables measured during the interview are defined as follows: (*a*) *Pt.'s Units*, the number of times the patient acted; (*b*) *Pt.'s Action*, the average duration of the patient's actions; (*c*) *Pt.'s Silence*, the average duration of the patient's silences; (*d*) *Pt.'s Tempo*, the average duration of each action plus its following inaction, as a single measure; (*e*) *Pt.'s Activity*, the average duration of each action minus its following inaction, as a single measure; (*f*) *Pt.'s Adjustment*, the durations of the patient's interruptions minus the durations of his failures to respond, divided by Pt.'s Units; (*g*) *Dr.'s Adjustment*, the durations of the interviewer's interruptions minus the durations of his failures to respond, divided by Pt.'s Units; (*h*) *Initiative*, the per cent of times, out of the available number of opportunities (usually 12) in Period 2, in which the patient acted again (within a 15-second limit) following his own last action; (*i*) *Dominance*, the number of times in Period 4 that the patient "talked down" the doctor minus the number of times the doctor "talked down" the patient, divided by the number of Pt.'s Units in that period; (*j*) *Pt.'s Synchronization*, the number of times the patient interrupted or failed to respond to the doctor, divided by the number of Pt.'s Units; (*k*) *Quickness*, the average length of time in Period 2 that the patient waited before taking the initiative following his own last action; (*l*) *Number of Interruptions*, the number of times the patient interrupted the doctor; (*m*) *Length of Interview*, the duration of the interview in minutes.

The above-defined variables were found to be characteristic of a given patient, in that they showed high reliability (test-retest stability) when he was interviewed by two different doctors a few minutes apart on the same day (7, 12) and when he was interviewed twice by one doctor over a seven-day interval (13). In studying the reliability of the input of the observer who records the raw data from which the variables are printed by the Interaction Chronograph, it was found that the observer's task is a highly reliable, primarily mechanical one, and that neither his memory nor a constant response-set is a probable explanation of the high reliabilities of the interviewee interaction variables (10). Although the measurements of a given patient's behavior made on one occasion were highly reproducible when he was re-interviewed, the values of each interaction variable differed widely from one patient to another. In other words, there were

[1] This investigation was supported by a research grant (M-1107) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

[2] The data were collected while the authors were at the Massachusetts General Hospital and Harvard Medical School.

large interindividual differences but small intra-individual variability in interviewee interaction patterns. The purpose of the present study was to explore what meaning could be attributed to these individual differences in terms of concurrent validity, i.e., the correlation between interview response behavior and other response behavior. This study represents a first approach, designed to provide leads for further investigation of the usefulness of the Interaction Chronograph in making meaningful differentiations among people or in predicting other kinds of behavior.

In an earlier paper, we reported some preliminary work on the relationship of the interaction variables to such organismic variables as age and sex (9). In the present study, we chose to explore these relationships further and to ascertain also the relationship of interview behavior to certain psychological test measures which are currently in clinical use, especially to those test variables which, on the surface, would appear to be measuring similar aspects of interpersonal behavior. For example, we believed that certain characteristics of an individual's verbal interview behavior could well be related to Wechsler Verbal IQ or to socioeconomic level, inasmuch as higher educational background and higher occupational status presumably are associated with the acquisition and continued use of certain verbal and social skills. We also expected that Rorschach variables might correlate with some aspects of interview behavior as measured by the Interaction Chronograph. Rorschach reaction-time, for instance, presumed to reflect habitual speed of response or, in Beck's words, "lack of inhibition" (2, p. 52), might be positively correlated with length of delay before making a verbal or other communicative response to the interviewer; and Rorschach Total Time and R (total number of Rorschach responses) might be related to the amount of interview time in which the interviewee was speaking. In other words, both situations (Rorschach and interview) might be found to measure, among other things, a characteristic readiness or quickness to respond, and a characteristic level of activity or productivity.

Selection of the test variables was of course based upon such hunches or expectations as those mentioned above. There was, however, no systematic attempt to construct a series of hypotheses since we did not want to feel committed to a particular theoretical viewpoint when our knowledge about the "meaning" of Interaction Chronograph variables is still sketchy and tenuous. Rather, it was felt that an exploratory study was all that could be attempted in the preliminary stages of validating these interviewee interaction patterns, and that any correlations which were found in the initial study of 20 *Ss* could serve as hypotheses for a cross-validation or replication study with a second group of 20 *Ss*.

METHOD

The 40 *Ss* in the original and cross-validation group were 26 outpatients and 14 inpatients (all white) of the Massachusetts General Hospital. The presenting problems of the patients, and the procedure of the interview were similar to those of other populations studied earlier (7, 12, 13). There were 25 females and 15 males ranging in age from 15 to 80, the mean age being 40. The same psychiatrist conducted all interviews, and the same observer recorded them from the other side of a large one-way mirror. A microphone was hung from the ceiling, and the voices were amplified and conducted into earphones worn by the observer. Facilities for visual and auditory observation were excellent.

Fifty patients, altogether, were interviewed in the order of their referral to the interviewer for psychiatric evaluation, their selection being without known bias. After the interview, the patients were referred for psychological testing. The psychologist had no knowledge of the patient's interview interaction behavior in order to insure independence of the psychological test measures and interview observations. Of the 50 referrals, 10 had to be eliminated from the experimental series because of refusal to take the psychological tests, borderline intelligence, or severe speech handicap. The tests were administered in one sitting, in uniform order, and with uniform instructions. First, educational and occupational data were obtained for use in determining socioeconomic level on the Hollingshead Index of Social Position. On this scale (4) the patients ranged from Class 1 (high executives of large concerns, proprietors, major professionals, etc.) to Class 5 (unskilled employees, welfare recipients, etc.). The mean for the population was Class 4 (skilled and semi-skilled workers, etc.). The complete Wechsler Adult Intelligence Scale was administered (16), followed by the simple language version of the Taylor Manifest Anxiety Scale (14, p. 288) and the Saslow Psychosomatic Screening Inventory (11). The next instrument administered was the Rorschach. Lest the amount of examiner-patient verbal interaction vary and possibly affect the latter's productivity, the examiner gave standardized instructions

and began the Rorschach without further comment.[3] If the patient rejected the first card or gave only one response, he was reminded to "be sure to report *everything* you see." Subsequent rejections were accepted without comment and questions were answered merely by a nod of the head or a word or two if necessary. This manner of presenting the Rorschach may have resulted in a smaller number of responses by some patients than would have been obtained with more encouragement. However, it was felt necessary to keep stimulus conditions, i.e., the Rorschach examiner's behavior, as similar as possible from patient to patient in order to make the Rorschach situation as standardized as were the interview conditions. The mean Rorschach total time of 8.23 minutes (shown in Table 1) was somewhat low, probably due in part to the manner of presentation and the nature of the population, but primarily to the fact that in this study total time was calculated as the time during which the patient was actually responding to the cards rather than as the total duration of the test. Last of all, the patient was asked to fill out the Allport Ascendance-Submission Scale. It was occasionally necessary, with those patients having the lower IQs, to help interpret items. It can be seen from the group mean for this test in Table 1 that the 40 patient *S*s obtained a score (−15.95), which presumably indicates greater submissiveness than the average college student, who obtains a mean score of approximately 0.

In order to control for interviewer (and psychological test examiner) experience differences, an odd-even method was used for selecting the 20 *S*s of the original study and the 20 *S*s of the cross-validational sample. Group 1 consisted of patients 1, 3, 5, etc., and Group 2 of the alternate patients. Statistical procedures were subsequently performed on each group separately, and then on both groups together.

RESULTS

Table 1 contains the means, standard deviations, and ranges of the psychological test and organismic variables for the total sample. There were no statistically significant differences between the two groups for these (test and organismic) variables. Means, etc., for the interview measures were comparable to those of groups previously studied (7, 12, 13). Table 2 contains the correlations (Pearson *r*) between the 14 interview variables on the one hand, and 13 psychological test measures, plus age, sex, and Hollingshead Index of socioeconomic status, on the other. The original matrix of intercorrelations contained 35 organismic and psychological test variables, but 19 are not presented here because they were found to be unrelated to any of the interview variables and would consume an inordinate amount of space. The excluded measures consisted of the Allport Ascendance-Submission Scale (which, developed on a college population, may have been merely unsuitable to the present clinic population) and a number of Rorschach scoring determinants. The Beck system was used for scoring Rorschach protocols, the only new classification used being "number of categories." This referred to the number of different Rorschach content categories utilized by a particular patient and was included in the present study because the distribution of scores among Beck's separate content categories was especially diffuse, with low frequencies in any single content classification making inappropriate the application of conventional correlation techniques. The distributions of the other variables differed somewhat in the degree to which they fulfilled the assumptions of the Pearson *r* but, in general, did not deviate markedly from a normal distribution. It was felt that, in view of the above as well as the Norton findings quoted in Lindquist (6), little precision would be lost in using Pearson *r*'s on the data in their original form. This is es-

TABLE 1

MEANS, STANDARD DEVIATIONS AND RANGES OF THE PSYCHOLOGICAL TEST AND ORGANISMIC VARIABLES FOR THE COMBINED SAMPLE OF 40 SUBJECTS

Variable	Mean	*SD*	Range
Age	40.33	13.93	15 to 80
Total IQ	97.65	12.70	77 to 129
Verbal IQ	101.45	13.40	77 to 137
Performance IQ	93.25	13.45	70 to 127
Taylor Anxiety	26.60	11.02	6 to 45
Psychosom. Inv.	13.50	10.14	0 to 44
Ascend.-Submiss.	−15.95	19.40	−47 to +39
Socioecon.			
(raw score)	58.40	13.92	11 to 77
(class)	4.12	1.12	1 to 5
Rorschach:			
Reaction time	17.98″	14.05″	4″ to 85″
Total time	8.23′	4.50′	1′ to 20′
No. Responses (R)	16.43	10.19	2 to 51
A + Ad%	56.18	16.56	24 to 99
No. of Categ.	6.58	4.48	2 to 13
% Popular	33.75	22.50	0 to 99
No. of Card Rejects	1.45	2.10	0 to 8

[3] The exact instructions were as follows: "I am going to show you some cards, and I want you to tell me whatever you see on them. This can involve the whole card or any part of it. Most people see a number of things on each card, so look carefully for everything you see represented there. When you are sure you have told me everything you see, put the card down and take the next one."

TABLE 2

PSYCHOLOGICAL TEST CORRELATES OF INTERVIEW INTERACTION BEHAVIORS FOR TWO GROUPS OF 20 *Ss* EACH AND COMBINED TOTAL GROUP OF 40 *Ss*

	Pt.'s Units			Pt.'s Action			Pt.'s Silence			Pt.'s Tempo		
	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.
Sex												
Age							−.56	−.45	−.49			
Total IQ	−.24	−.45	−.34				.27	.23	.26			
Verbal IQ	−.36	−.40	−.36	.20	.21	.20	.21	.29	.26	.22	.28	.25
Perf. IQ												
Anxiety							−.23	−.20	−.20			
Psychosom.	−.38	−.41	−.39	.29	.50	.39				.29	.52	.40
Socio-Econ.							−.20	−.31	−.22			
Ave. RT	−.39	−.08	−.25	.60	.20	.44				.60	.17	.42
Tot. Time							−.27	−.23	−.25			
No. Responses												
D%	.39	.11	.25	−.42	−.26	−.35				−.43	−.30	−.36
Dd%	−.60	−.37	−.47	.65	.25	.48				.65	.29	.48
A + Ad%	.41	.24	.31	−.45	−.13	−.29				−.43	−.14	−.28
% Popular												
No. Categ.												

	Pt.'s Activity			Pt.'s Adjustmt.			Dr.'s Adjustmt.			Pt.'s Synchron.			Dr.'s Units		
	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.
Sex															
Age				.24	.30	.27	.31	.31	.30				−.64	−.11	−.31
Total IQ							−.37	−.41	−.36				−.25	−.44	−.33
Verbal IQ	.18	.24	.21				−.42	−.35	−.34				−.38	−.38	−.36
Perf. IQ							−.23	−.41	−.31						
Anxiety										−.28	−.29	−.28			
Psychosom.	.29	.52	.40							−.36	−.43	−.39	−.34	−.43	−.38
Socio-Econ.							.31	.26	.23						
Ave. RT	.60	.16	.44							−.58	−.12	−.38			
Tot. Time															
No. Responses															
D%	−.42	−.31	−.37							.45	.28	.37	.46	.16	.31
Dd%	.65	.32	.51							−.53	−.31	−.43	−.61	−.37	−.48
A + Ad%	−.46	−.17	−.32							.47	.21	.33	.46	.25	.34
% Popular															
No. Categ.															

	Pt.'s Initiative			Pt.'s Quickness			Pt.'s Dominance			Pt.'s No. of Interrup.			Length Interview		
	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.	1	2	Tot.
Sex							.41*	.58*	.50*	.04*	.63*	.36*			
Age	.50	.45	.47	.48	.45	.47									
Total IQ							−.11	−.54	−.33	−.18	−.41	−.32	.13	.33	.24
Verbal IQ							−.23	−.61	−.42	−.29	−.45	−.38	.28	.27	.27
Perf. IQ															
Anxiety	.57	.36	.44	.51	.25	.37									
Psychosom.	.28	.28	.27										.25	.52	.39
Socioecon.															
Ave. RT										−.33	−.16	−.21	.59	.20	.40
Tot. Time	.35	.24	.31	.25	.36	.31	−.26	−.33	−.27	−.25	−.22	−.20			
No. of Responses							−.36	−.43	−.37	−.22	−.21	−.18			
D%							.25	.23	.23	.28	.28	.24	−.34	−.38	−.33
Dd%							−.28	−.27	−.25	−.35	−.26	−.24	.54	.26	.37
A + Ad%				−.32	−.27	−.29							−.50	−.07	−.26
% Popular				−.41	−.15	−.29	.37	.39	.37						
No. of Categ.							−.25	−.38	−.33						

	$N = 20$	$N = 40$
Note:—r required for significance at .05 level is	.44	.31.
r required for significance at .01 level is	.56	.40.

pecially true since, in previous studies of interview behavior (9), comparison of Pearson *r*'s and rho's have yielded negligible discrepancies in most cases.

Table 2 contains only those correlations which are significant ($p = .01$ or .05) in both the original and cross-validational groups or which indicate for both groups a consistent trend approximating significance. Thus, Table 2 includes only those correlations which appear in group one and are replicated in the second group of 20 since, in a matrix of 35 by 14 variables, one would expect to find a number of misleading "significant" correlations based on chance alone.[4]

While all fourteen interview variables are presented in Table 2, the present paper will focus on only five major variables because most of the others are some kind of combination or derivation of these. The first and second of the five variables to be discussed are Pt.'s Units, (the number of times the patient talks or communicates by gesture during the standardized interview) and Pt.'s Action (the average duration of his utterances). As reported in a previous paper (9), Pt.'s Units has a high negative correlation with Pt.'s Action. (In the present total sample, this correlation is −.85)[5]. In Table 2 it can be seen that the correlation coefficients between the test variables and the interaction variables also seem to define these two interview characteristics as opposites, and thus begin to give further definition to them. For example, in the two samples combined, the negative *r*'s between Pt.'s Units and Total IQ (−.34) and Verbal IQ (−.36) indicate that many short interview units are significantly associated with lower Total and Verbal IQ, while fewer utterances are significantly associated with relatively higher Total and Verbal IQ. The positive (but not statistically significant) *r*'s of .20 and .21 between Pt.'s Action and Verbal IQ may indicate a tendency in both groups for the longer utterances to be found in those with the higher Verbal IQ. Taking these observations together, it would appear that during the interview the more intelligent patients tend to speak a smaller number of times but with longer durations per utterance, while the less intelligent patients speak a greater number of times but with a shorter duration each time.

Table 2 presents other psychological test correlates of the number of times (units) a patient interacts during the standardized interview. Thus the *r* of −.25 between Pt.'s Units and Average Reaction Time; of .25 between Pt.'s Units and D%; of −.47 between Pt.'s Units and Dd%; and of .31 between Pt.'s Units and A + Ad%, describe the (less intelligent) patient with many short interview units as responding more quickly to the Rorschach cards, tending to give more large detail responses but fewer small detail and more animal responses. In parallel fashion, the correlation coefficients between Pt.'s Action and the same Rorschach variables portray the (more intelligent) patient with the longer interview utterances (and consequently a smaller number of them) as responding more slowly to each Rorschach card, tending to give fewer large detail but more small detail responses, and fewer animal responses. However, intelligence level alone does not account for the above observations since, in the two samples studied individually or combined, the correlation coefficients between IQ and D% and IQ and Dd% were not statistically significant. (These coefficients are omitted from Table 2).

The *r*'s of −.39 between Pt.'s Units and Psychosom. (the number of symptoms checked by the patient on the Saslow Screening Inventory) and .39 between Pt.'s Action and Psychosom. indicate that the patient who talks a greater number of times but with short utterances in the interview reports fewer psychosomatic symptoms in relation to stress in his daily life. (There was no significant association between IQ and number of symptoms checked on the inventory). The explanation of these *r*'s of −.39 and .39 is unclear. Perhaps they reflect denial, caution, or repression on the part of these patients of relatively lower IQ, who tend also to give more stereotyped Rorschach responses and who speak many times in the interview because, under the conditions of the

[4] It is clear in Table 2 that many of the correlation coefficients between interview and psychological test variables, while reaching the .05 or .01 levels of statistical significance are, nevertheless, of a low order of magnitude. Taken together, however, they seem to contribute to the understanding of interview behavior.

[5] This is partly a function of the partially standardized interview. That is, three of the subperiods are fixed in duration, and thus the patient who speaks in long utterances will not have the opportunity to use a large number of units and conversely, the individual with a relatively large number of units must have spoken in fairly short utterances.

standardized interview, each of their utterances is brief.

Certain other significant correlates of Pt.'s Units and Pt.'s Action (not shown in Table 2) are of interest. These are *r*'s of −.33 between Pt.'s Units and Initiative; of .34 between Pt.'s Units and Pt.'s Dominance; and of .56 between Pt.'s Units and Pt.'s No. of Interruptions. The corresponding *r*'s with Pt.'s Action are .29, −.38, and —.62, respectively. These coefficients portray the patient who talks often during the interview as taking less initiative in the "silence stress" subperiod, and as interrupting and talking down the interviewer more often during the "interruption stress" subperiod than the patient who speaks a smaller number of times but in longer units.

In summary, the person with a large number of Pt.'s Units and the person with few units and correspondingly longer durations of utterance per unit present almost opposite pictures. In Table 2, there is a consistent though nonsignificant trend in both groups for the person with longer utterances to be more intelligent (*r* of .20 between Pt.'s Action and Verbal IQ). He responds more slowly to the Rorschach (*r* of .44 between Pt.'s Action and Average RT), but apparently with more individually meaningful material (*r* of .48 with Dd%) in the sense that he reports seeing more rare details, and he gives fewer of the relatively easy large detail (*r* of −.35 with D%) and animal responses (*r* of −.29 with A + Ad%). He also reports having more psychosomatic symptoms, as reflected in the Saslow Screening Inventory (*r* of .39 with Psychosom.) and for this reason might be thought of as being less guarded in his test-taking behavior. In line with this hypothesis, he is again the opposite of the "many units but each of short duration" person in that he tends to show more initiative (*r* of .29 between Pt.'s Action and Pt.'s Initiative, not shown in Table 2) in the ambiguous part of the interview when the interviewer remains silent by consistently failing to respond to the patient. Finally, the more intelligent, "few units but each of long duration," individual lacks the less intelligent "many units but each of short duration" individual's tendency to dominate and interrupt (*r* of −.38 between Pt.'s Action and Pt.'s Dominance, and of −.62 with Number of Interruptions). Pt.'s Initiative and Pt.'s Dominance, themselves, are not significantly intercorrelated. These figures are not shown in Table 2.

It is interesting to note in Table 2 that the average duration of Pt.'s Silence is *not* the opposite of Pt.'s Action in regard to the psychological test correlates. This finding is further substantiated by a recently completed factor-analysis in which it was found that these two interview variables (average Pt.'s Action and average Pt.'s Silence) are almost wholly independent factors. In fact, the results in Table 2 indicate that intelligence (both Total and Verbal) is positively related to Pt.'s Silence just as to Pt.'s Action. The relationship falls just short of statistical significance, but is consistent in both groups between Verbal IQ and both Pt.'s Action and Pt.'s Silence, and additionally between Total IQ and Pt.'s Silence. Another consistent trend is for Pt.'s Silence to be positively correlated with higher socioeconomic level. (The Hollingshead Index ranges from a high score for the lowest socioeconomic group to a low score for the highest socioeconomic group, thus the negative correlation.) Pt.'s Silence is also negatively associated with Taylor Anxiety score. It thus would appear that silence behavior in the interview has a low positive relationship to intelligence and socioeconomic level and is negatively related to anxiety, with the longer average interview silences tending to be found in patients of higher intelligence, higher socioeconomic level, and lower Taylor anxiety levels. In Table 2 it can be seen further that the patient's age is the variable having the greatest relationship to Pt.'s Silence (*r* of −.49), indicating that younger *S*s had the longer average interview silences.

On the basis of these results, one might wonder if Pt.'s Silence behavior has two aspects: one, a degree of tolerance for silence among patients with lower anxiety scores and also in patients of higher socioeconomic status and higher IQ; the other, a factor associated with age (poise, self-assurance, or lack of defensiveness) and resulting in longer waiting periods before taking the initiative in the silence stress subperiod among the younger *S*s. The former culturally "positive" aspects of Silence behavior account, however, for only a small proportion of the total variance, the correlations being consistent in both the original and cross-validational samples but not attaining statistical significance. The latter "negative" and

larger aspect of Pt.'s Silence behavior seems to be a lack of ability in the younger *S*s to respond as quickly (with as short silences) as the older *S*s in the silence stress subperiod of the interviews. In this regard it should be mentioned (data not presented in Table 2) that Pt.'s Silence and Pt.'s Initiative have a correlation of −.69 in this sample, which indicates that whatever accounts for the Pt.'s Silence variable (an average duration measure) partly accounts also for the *S*'s failure to take the initiative after a double silence in Period 2 (a frequency measure).

Turning to the variable, Pt.'s Initiative, also shown in Table 2, it can be seen that age has both a correlation of −.49 with Pt.'s Silence (the older the patient the shorter his average interview silences) and +.47 with Pt.'s Initiative (the older the patient the more often he took the initiative during the silence period), thus supporting the possibility of the presence of a "poise" factor, increasing with age. Further suggestion of the plausibility of this "poise" interpretation is given by a recent analysis of data from a previously studied population of 60 out-patients (9). These results are shown in Table 3. For these 60 *S*s, the mean initiative score of those under age 30 was .65 during the first interview, while the mean initiative score of those *S*s over 30 was .78 (*F* test significant at .01 level). The value of the correlation ratio, epsilon, was .34 (significant at .01 level), indicating that this relationship between age and increasing interview initiative is a moderate one. All of the *S*s had a second interview, at which time the age-initiative relationship disappeared; mean initiative scores for the younger and older groups were .70 and .73 respectively, with the *F* test between age and initiative dropping to an insignificant .39 and epsilon to zero. This finding of less initiative among the younger group in the first interview is cross-validated in the present sample of 40 *S*s. The epsilon value of .46, shown in Table 3, corresponds to the Pearson *r* of .47 (for the same data) shown in Table 2.[6]

[6] Twenty of this group have been re-interviewed after an 8-month interval. For these 20, the age-initiative relationship again disappears in the second interview. The mean initiative score for the under 30 group is .66, while for the over 30 group it is .72. The *F* test of .67 is not significant; nor are the *r* of −.09 and the *e* of .00. The same 20 *S*s showed a positive (*r* of .26) age-initiative relationship in their first interviews.

TABLE 3
PATIENTS' AGE AND INITIATIVE DURING INTERVIEW

Mean Initiative Score	Patient's Age		*F* test	*e*
	Under 30	Over 30		
Present sample, $N = 40$				
a. First interviews	.56	.77	11.88**	.46**
Previous sample, $N = 60$				
a. First interviews	.65	.78	7.71**	.34**
b. Second interviews	.70	.73	.39	.00

** Significant at the .01 level of probability.

Given the hypothesis of a "poise" or "set" factor, developing with age, it appears to operate largely in the first interview and apparently produces no behavioral effect when the patient has some familiarity with the stimulus situation, as in the second or retest interview. This hypothesis has certain face validity when one thinks of either therapeutic or social contacts with adolescents, who seem to have considerable reluctance to initiate with adults but who manage to respond at some length when the initiative has been taken by someone else. At any rate, it is interesting that age is related to Pt.'s Initiative and Pt.'s Silence, but not to Pt.'s Action.

Other psychological test correlates of Pt.'s Initiative, presented in Table 2, further suggest that it is, in some ways, the opposite of the Pt.'s Silence variable. That is, Pt.'s Initiative increases positively with anxiety score and number of psychosomatic symptoms, just as Pt.'s Silence is negatively related to anxiety. Thus, one could speculate that Pt.'s Initiative is, in part, a function of the magnitude of his anxiety level as the latter is assessed by the Taylor Anxiety scale; the more anxious *S*s show the higher percentage of interview initiative (*r* of .44). That initiative may be a reflection of some kind of "drive" ("anxiety") is further suggested by the fact that it seems to be related to effort expended on the Rorschach, since it is significantly and positively correlated ($r = .31$) with total time spent in responding to the ink blots. Possibly the percentage of times an *S* takes the initiative under conditions of planned interview silence may be correlated with a variety of laboratory learning tasks, as in the Taylor "anxiety" level.

In Table 2 it can be seen that Pt.'s Domi-

TABLE 4
PATIENTS' SEX AND DOMINANCE-SUBMISSION PATTERN DURING INTERVIEW

Mean Dominance Score	Females	Males	*F*-test	ϵ
Present sample, $N = 40$				
a. First interviews	−.42	−.71	8.44**	.50**
Previous sample, $N = 60$				
a. First interviews	−.33	−.49	5.65*	.26*
b. Second interviews	−.26	−.53	12.50***	.40**

* Significant at the .05 level of probability.
** Significant at the .01 level of probability.
*** Significant at the .001 level of probability.

nance seems to be the only variable which is significantly related to sex, the female patients being relatively less submissive (or more dominant) than males with the male interviewer. Since this dichotomous variable (sex) was not suited to Pearson *r*'s, *F* tests and epsilon correlations were obtained and are the values shown (.41, .58, and .50) across the sex row in Table 2. The values of epsilon needed for significance are somewhat lower than those for *r*, being .39 and .53 for the .05 and .01 levels, respectively, with an *N* of 20, and .27 and .37 with an *N* of 40.

A further analysis of the sex-dominance relationship is presented in Table 4. Mean Pt.'s Dominance score for females in the present combined sample of 40 *S*s was found to be −.42 and for males, −.71. The *F* value for this difference was 8.44, significant at the .01 level, as was the epsilon correlation ratio of .50. Since this finding of significantly less submissiveness in the female patients relative to the male patients was contradictory to expectations, built upon cultural stereotypes of the dominant male, one may wonder if the factor of illness is not an important one here. Perhaps one correlate of these persons' psychiatric difficulties is that they do not fit into their expected cultural roles. We hope to repeat the sex-dominance analysis on a "normal" group to test the validity of this hypothesis. However, the finding does appear to be consistent among patient groups since the same relationship was obtained in a previously studied sample (9). As shown in Table 4, in that sample of 60 outpatients, the epsilon correlation value was .26, significant at the .05 level of confidence for the first interviews. For the second interviews, the epsilon correlation increased to .40, significant at the .01 level. The *F* tests similarly indicate an increase in this relationship for the second interview. Thus, in both the first and retest interviews, the males were found to be more submissive (were dominated more) than were the females. However, the negative mean dominance scores indicate that both male and female patients, as groups, are more submissive than dominant, with male patients being more so. Since we have not yet studied a group of normal subjects of comparable characteristics, we do not know if this difference exists only in a patient population and if so, whether women patients are relatively more dominant than the normal members of their sex, whether the males are more submissive than the normal members of their sex, or both.

Returning to Table 2, we see further that Pt.'s Dominance was negatively associated with Total and Verbal IQ, i.e., the less intelligent the *S* the more dominance and interruption he showed in the interview. Looking to the Rorschach correlates of the dominance variable, we find a picture of stereotypy or guardedness. Patients high in Pt.'s Dominance tended to spend less total time with the ink blots, gave fewer responses, reported significantly more large detail responses, perceived more populars, attended to fewer small details, and used a smaller number of content categories. Thus, the patient who shows relatively more interview dominance tends to be less intelligent, displays a relative lack of imagination on the Rorschach test, and apparently either has less ego-involvement in his Rorschach performance or is significantly more guarded about his responses, returning cards relatively quickly and after fewer responses. He spends less total time in Rorschach performance, although his reaction time does not differ from those who are less dominant, so that he apparently does not respond hastily and without thinking but simply differs in the quantity (minus *r* with Total Number of Rorschach responses) and quality (plus *r* with P%, D%, and minus *r* with No. Categ.) of his productions. This relative stereotypy or rigidity of the high interview Dominance individual is another interesting finding which merits further exploration.

It may be noted that throughout the analysis the correlations between several Rorschach variables and the Interaction Chronograph

variables are lower in Group 2 than in Group 1. Inspection of the sigmas of these variables (not shown separately in Table 1) revealed that, in the case of reaction-time, the sigma of Group 2 was less than half that of Group 1, while for D% and Dd% the sigmas were ⅔ the size of those for Group 1. Since the size of the Pearson *r* is dependent upon variability within the group or is lowered by a restricted range of scores, the smaller standard deviations may account for the lower correlations in Group 2. The variable A + Ad%, however, (which also had a lower correlation with the interview variables in Group 2) did not present a similar difference in range and sigma between the two groups, and this factor therefore does not explain the discrepancy in the case of A + Ad%.

DISCUSSION

The present observations suggest the possibility of differentiating among individual patients along several definable dimensions. There seem to be some consistent organismic and psychological test differences between people who speak in long and short utterances in the partially standardized interview, between those who have short or long silences, between those who show much or little interview "initiative" in the silence stress subperiod, and between those who "submit to" or who "dominate" the interviewer when the latter has interrupted them in the interruption stress subperiod.

Intelligence was found to be associated with certain interview interaction patterns, although socioeconomic level was not. It should be noted, however, that in our combined clinic and inpatient population, we were not able to obtain an adequate sampling of the higher socioeconomic levels and do not feel that we have explored adequately the possibility of a relationship between socioeconomic level and interview interaction pattern.

In concurrence with the negative results of several other recent investigators, in this study Rorschach perceptual determinants were not found to correlate with interview behavioral characteristics, although a few Rorschach content measures proved to be more promising. This failure of such Rorschach variables as Movement, Color, etc., to correlate with interview behavior is in some ways similar to the recent findings of Knopf (5), who reported that a few Rorschach content categories were useful in differentiating among widely differing diagnostic groups, while Rorschach perceptual variables failed even to differentiate reliably a schizophrenic group from other diagnostic groups.

While the present investigation was designed to study the possible psychological test correlates of interview interaction behavior, and thus to give meaning to individual differences in the interview, it may be observed that the interaction variables may themselves be thought of as contributing meaning to the psychological test variables with which they are significantly associated. That is, the standardized interview could equally well have been chosen as the criterion against which to validate the psychological tests. For example, we have shown that those *S*s of higher intelligence have different interview characteristics than those of lower intelligence, since they speak in longer utterances and tend to interrupt or dominate the interviewer less frequently. Similarly, it is interesting that those *S*s who give stereotyped Rorschach behavior in terms of fewer responses, more large details, more populars, fewer content categories, etc., tend to dominate the interviewer and can, in that respect, be said to show somewhat less social sensitivity in their interview interaction. Also, those who score relatively high on the Taylor Anxiety scale are less silent and take the initiative more frequently than those who are presumably less anxious. A parallel could be drawn between this and some of the recent literature indicating that those with higher anxiety scores tend to obtain higher learning scores on conditioning or easy learning tasks.

The present study, in any event, is to be considered only an initial approach to an understanding of the validity or "meaning" of the interview behaviors measured by the Interaction Chronograph. It has been an exploratory approach through which we hoped to gain leads for more systematic research. The examination of psychological test correlates has been especially rewarding when one considers that in the present study we have related them only to the simplest, discrete interview variables, and so far have not attempted to relate them to patterns of these variables or to sequential changes in behavior as the interview proceeds.

The relationship of the psychological test and other variables to more complex interview interrelationships and to sequential variations in interviewee behavior demands systematic inquiry.

SUMMARY

The present study was an attempt to examine the "meaning" of Interaction Chronograph patterns through concurrent validity, relating interview interaction patterns and psychological test behavior. A number of meaningful correlations was found in a sample of 20 patient *Ss* and cross-validated in a second similar sample. Individuals who speak in short utterances during the interview are less intelligent and seem to be more "on their guard" or more stereotyped in their psychological test productions. Those with long utterances are somewhat more intelligent and are less cautious or denying, giving less conventional performance on the Rorschach test and reporting more psychosomatic symptoms. Initiative during the silence period of the interview is hypothesized to be related to "poise" or self-assurance (or some other "set" factor) since it seems, in the first of a pair of interviews, to be greater with age. The individual with a high Pt.'s Dominance score is of relatively lower intelligence and presents a Rorschach picture of a smaller number of responses, high P% and A + Ad%, a smaller number of content categories, etc. This was interpreted as indicative of stereotypy or rigidity. Pt.'s Dominance seems also to be the only interview variable which is related to sex, female patients being relatively less submissive than male patients. The present investigation can also be thought of as providing certain behavioral (interview) correlates of psychological tests (Wechsler, Rorschach, etc.) currently in clinical use.

REFERENCES

1. ALLPORT, G. W., & ALLPORT, F. H. *The A-S Reaction Study: A scale for measuring ascendance-submission in personality.* (2nd ed.) Boston: Houghton Mifflin, 1939.
2. BECK, S. J. *Rorschach's test—a variety of personality pictures.* Vol. 2, New York: Grune & Stratton, 1949.
3. CHAPPLE, E. D. The Interaction Chronograph; its evolution and present application. *Personnel*, 1949, **25**, 295–307.
4. HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class & mental illness.* New York: Wiley, in press.
5. KNOPF, I. J. Rorschach summary scores in differential diagnosis. *J. consult. Psychol.*, 1956, **20**, 99–104.
6. LINDQUIST, E. F. *Design and analysis of experiments in psychology & education.* Cambridge: Riverside Press, 1953.
7. MATARAZZO, J. D., SASLOW, G., & GUZE, S. B. Stability of interaction patterns during interviews: a replication. *J. consult. Psychol.*, 1956, **20**, 267–274.
8. MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH G. The Interaction Chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347–367.
9. MATARAZZO, J. D., SASLOW, G., MATARAZZO, RUTH G., & PHILLIPS, JEANNE S. Stability and modifiability of personality patterns during a standardized interview. In P. A. Hoch and J. Zubin (Eds.), *Psychopathology of communication.* New York: Grune & Stratton, 1957.
10. PHILLIPS, JEANNE, S., MATARAZZO, J. D., MATARAZZO, RUTH G., & SASLOW, G. Observer reliability of interaction patterns during interviews. *J. consult. Psychol.*, 1957, **21**, 269–275.
11. SASLOW, G., COUNTS, R., & DUBOIS, P. H. Evaluation of a new psychiatric screening test. *Psychosom. Med.*, 1951, **13**, 242–253.
12. SASLOW, G., MATARAZZO, J. D., & GUZE, S. B. The stability of Interaction Chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, **19**, 417–430.
13. SASLOW, G., MATARAZZO, J. D., PHILLIPS, JEANNE S., & MATARAZZO, RUTH G. Test-retest stability of interaction patterns during interviews conducted one week apart. *J. abnorm. soc. Psychol.*, 1957, **54**, 295–302.
14. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
15. ULETT, G. A., GLESER, GOLDINE, STARR, P., HADDOCK, J., LINGLEY, L., & LAWLER, ANN. Psychiatric screening of flying personnel: Further studies toward the development of an electroencephalographic screening technique. *USAF Sch. of Aviat. Med.*, 1953, Project No. 21-0202-0007 (Report No. 5).
16. WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corp., 1955.

Received May 28, 1957.

INTERRELATIONSHIPS AMONG INDIVIDUAL DIFFERENCE VARIABLES, BEHAVIOR IN PSYCHOTHERAPY, AND VERBAL CONDITIONING[1,2]

IRWIN G. SARASON

University of Washington

THE present experiment dealt with the interrelationships among three variables: (*a*) *S*'s score on the scales of a personality inventory, (*b*) therapists' ratings of *S*'s behavior in the psychotherapy situation, and (*c*) *S*'s performance in a verbal conditioning situation.

Within recent years there has been an increase in the frequency of studies on verbal conditioning (1, 4, 9, 13, 14). These studies have centered on the effects of *E*'s reinforcement of *S*'s verbal responses (of a predetermined class) on *S*'s subsequent verbal behavior. One is struck by the similarity of this two-person verbal conditioning situation to psychotherapy and everyday conversational situations. In both of these kinds of interpersonal interactions, individuals emit responses that may or may not be reinforced by another individual. For example, in the psychotherapy situation, the therapist may be particularly interested in reinforcing responses (verbalizations) of the shy withdrawn patient which indicate assertiveness and independence. On the other hand, with other kinds of patients and at certain stages of psychotherapy, the therapist may be inclined to reinforce responses of the patient which indicate dependence on or compliance with the therapist.

Implicit in this example is the importance of initial individual differences in *S*s' anticipation of reinforcement and their reaction to reinforcement. In a random sample of *S*s, one would not expect all *S*s to respond to reinforcement in the same way. Tatz (10), for instance, has recently demonstrated the significant role of individual differences among *S*s in a verbal conditioning situation in the extent to which they anticipate punishment or reward from the experimenter. The present study was concerned with the effect of certain personality characteristics on verbal conditioning.

The instrument employed to select *S*s was the Autobiographical Survey, a true-false test that yields six scores: (*a*) Test Anxiety (TA), (*b*) General Anxiety (GA), (*c*) Lack of Protection (LP), (*d*) Need for Achievement (NA), (*e*) Defensiveness (D), and (*f*) Hostility (H). Brief descriptions of the scales are presented below. While the names of five of the six scales fairly well suggest the nature of their content, it should be mentioned that the Lack of Protection scale consists of items all of which were written to adhere to Freud's definition of anxiety, i.e., feeling unprotected and unable to cope with novel, threatening situations.

Two predictions were made about the verbal conditioning performance of *S*s differing in scores on certain scales: 1. *S*s with high scores on the Test Anxiety (TA), General Anxiety (GA), and Lack of Protection (LP) scales condition faster than *S*s with low scores on these scales. These predictions were based on the hypothesis that in a novel situation such as verbal conditioning, in which *S* has no prior knowledge of what constitutes a correct response, anxious individuals, i.e., individuals who admit fear, tension, and inadequacy feelings in strange situations, are relatively more likely than other *S*s to increase the frequency of responses reinforced by *E*'s uttering "mm-hm." Too, individuals who feel inadequate in coping with strange situations may be more likely to accept direction (in this case, *E*'s reinforcement) from an authority figure. 2. Individuals with high scores on the Defensiveness and Hostility scales condition significantly more slowly than low scoring *S*s. This prediction was based on the hypothesis that people who are defensive or hostile in interpersonal situations may frequently be unlikely to accept interpretations, reinforcements, and cues from others.

At this point it is well to emphasize that the predictions made for the verbal conditioning

[1] This research was carried out at the West Haven, Connecticut VA Hospital. The writer is indebted to Jacob Levine, chief psychologist, for his cooperation and encouragement while the research was in progress. John Rakusin contributed many helpful suggestions, for which the writer is grateful.

[2] Portions of this paper were presented at the May, 1957, meeting of the Western Psychological Association.

behavior of *S*s are similar to a set of predictions made concerning the behavior of patients in psychotherapy. In carrying out this study, attention was focused on the possibility of studying complex variables of presumed importance in psychotherapy (e.g., the defensiveness of the patient) in simpler, more controllable situations such as verbal conditioning. With a better understanding of the determinants of behavior in such relatively simple interpersonal situations, it is possible that lines of experimental attack may be opened in more complex situations such as psychotherapy. In order to gain some preliminary information on this possibility, therapists' ratings of patients' behavior in therapy were obtained for the majority of *S*s in this experiment. Thus, it was hoped that some initial rough comparative statements might be made concerning the behavior of *S*s in both the therapeutic and verbal conditioning situations.

The *S*s were rated by their therapists on three characteristics which seemed to be related to certain characteristics of *S*'s Autobiographical Survey performance. These ratings concerned the degree of compliance with which the patient accepts therapist's interpretations, the degree of dependence the patient evinces towards the therapist, and the degree of hostility the patient displays towards the therapist. As with the H scale of the Survey, it was expected that *S*s rated by therapists as being hostile would condition more slowly than other *S*s. Similarly, high compliance and dependence therapist ratings were expected to be related to faster verbal conditioning on the basis of the need of the patient to please authority figures.

METHOD

Subjects

The *S*s were 60 neurotic and psychotic patients on the intensive treatment NP wards of the West Haven, Connecticut, VA Hospital. All but two of the *S*s were men, and the age range for the entire group was 22–52. All *S*s performed in the verbal conditioning situation and filled out the individually administered Autobiographical Survey.

Autobiographical Survey

This inventory consists of 180 statements to be answered True or False.[3] The first two items are neutral, nonscored items, and the remaining 178 are divided among the six scales. A brief description of each scale follows, together with examples of the kinds of items on the scale.

1. *Test Anxiety* (*TA*). This scale consists of 21 statements, most of which were taken from Sarason's Test Anxiety Questionnaire (6) and rewritten to fit the requirements of a true-false test rather than a graphic rating scale. Example: "While taking an important examination, I perspire a great deal."

2. *General Anxiety* (*GA*). The 24 items of this scale purport to tap the extent to which the individual experiences anxiety in a wide variety of situations other than testing. The statements used on this scale include some items to be found in the MMPI (5), the Taylor Anxiety Scale (11), and Sarason's General Anxiety Questionnaire (3). Example: "I worry about my social adjustment more than do most people."

3. *Lack of Protection* (*LP*). Some of the 42 items on this scale were taken from the scale presented by Mark (7). Most of them, however, were written by the author so as to adhere to the Freudian definition of anxiety: "The danger situation is the recognized, remembered and anticipated situation of helplessness. Anxiety is the original reaction to helplessness in the traumatic situation, which is later reproduced as a call for help in the danger situation" (2, p. 115) ". . . the situation of privation in which stimuli reach an unpleasurable magnitude and intensity without an ability to cope with them psychically and thus provide for their discharge, must represent to the infant a situation analogous to the birth experience, a repetition of the danger situation" (2, p. 76). Example: "When I was a child, I often wondered about how much my father loved me."

4. *Defensiveness* (*D*). Most of the 27 items of this scale were drawn from the K scale of the MMPI (5). This scale was included in order to obtain a measure of the individual's need to present himself in a favorable light. Example: "I am always patient with others."

5. *Hostility* (*H*). This scale, consisting of 23 items, taps the degree to which the individual admits to feelings of anger and hostility in a variety of situations. Many of these items were drawn from the MMPI. Example: "I rebel against bossy teachers."

6. *Need for Achievement* (*NA*). The 41 items of this scale, many drawn from the MMPI, were intended to measure the extent to which the individual feels called upon to attain a high level of performance in a variety of situations. Example: "My parents expected the highest level of achievement from their children."

Psychotherapists' Ratings

The therapists who made ratings of patients' behavior in psychotherapy sessions were chiefly

[3] A copy of the Autobiographical Survey and its scoring keys has been deposited with the American Documentation Institute. Order Document No. 5517, remitting $1.25 for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies.

psychiatric residents with one year's experience in psychotherapy. Some of the patients used in this study were rated by either staff clinical psychologists or advanced clinical psychology trainees, when these clinicians were the patients' therapists.

The therapists were asked to rate the three behaviors in question as occurring *almost always*, *frequently*, *seldom*, or *never* in psychotherapy. The statements rated by therapists in the experiment were: (*a*) How compliant is this patient to the therapist's requests and in accepting his interpretations? (*b*) How dependent is this patient on the therapist? (*c*) To what extent does this patient express hostility to his therapist?

Procedure

The relationship in time for each *S* between administration of the Survey and performance in the verbal conditioning situation was not rigidly controlled because of exigencies in hospital schedule and routine. However, in the great majority of cases, *S*s were not administered the Survey until after they had performed in the verbal conditioning situation.

All *S*s were given the following preliminary instructions: "I am going to show you these cards one at a time. For each card make up a two-word sentence using the pronoun at the top of the card and one of the three verbs at the bottom. You may choose any verb you like for each card."

Following these instructions, *S*s received 160 verbal conditioning trials. The *E* did not reinforce any responses made by *S* on Trials 1–20. Each trial consisted of the presentation to *S* of a 3 x 5 card, on the top of which was typed the word *he* and on the bottom of which were typed three verbs of equal frequency of occurrence according to the Thorndike-Lorge lists (12). One of the three verbs was a verb of the reinforced class; the other two were neutral words not of the reinforced class. The reinforced class and the general procedure used were the same as those used in a recent study by Sarason (8). Reference should be made to that study for a detailed description of the verbal conditioning procedure. Verbs of the class Verbal Activity were chosen as the reinforced class, and whenever *S* used a verb of this class, *E* uttered "mm-hm." Verbal Activity included verbs having to do with the use of the vocal apparatus, e.g., *talked*, *whispered*.

Following the conditioning series, a brief interview was conducted along the lines described by Sarason (8) in order to determine the level of *S*'s ability to verbalize the basis upon which *E* reinforced *S* and, indeed, if *S* was even aware of *E*'s reinforcements. None of the *S*s as judged by the interviews was aware of the principle on the basis of which *E* emitted "mm-hm." Seven *S*s, however, did comment that they had noticed that *E* often uttered "mm-hm."

RESULTS

Psychometric Findings

In order to assess meaningfully the influence on verbal conditioning of the individual differences tapped by the six personality scales of the Autobiographical Survey, it was of interest to obtain the intercorrelations as well as the reliabilities of the scales. Reliabilities, obtained by retesting 40 of the *S*s in this study two weeks after they initially took the Survey, were all above .70. Only the reliability of the NA scale fell below $r = .78$. The high interscale correlations in Table 1 suggest that the personality characteristics which the scales purport to measure are either defined in such a way as to be overlapping or are not "in nature" independent of each other. Whichever alternative is the case, however, in many of these correlations a large part of the variance in one variable is still unaccounted for by the variance in the other. For example, the TA and GA scales are correlated .46. This would suggest, for example, that although some *S*s admit to anxiety both in testing and a variety

TABLE 1
INTERCORRELATIONS AMONG SCALES OF AUTOBIOGRAPHICAL SURVEY
($N = 60$)

	GA	TA	LP	NA	D	H
GA		.46	.40	.24	−.61	.37
TA			.36	.39	−.54	.36
LP				.44	−.52	.56
NA					−.31	.41
D						−.61
H						

$r = .28 = .05$ level.
$r = .33 = .01$ level.

of other situations, many individuals admit to anxiety in testing but *not* in many everyday life situations. Such findings suggest that in addition to pursuing the problem from the viewpoint of clarifying the concepts and factorially purifying the scales of the Survey, an important aspect of study would be the differential characteristics of *S*s who score consistently high on anxiety scales such as TA, GA, and LP as compared with *S*s who do not display a consistent pattern.

Two sets of correlations of particular interest concern those of the Defensiveness and Hostility scales. With respect to the latter, one is struck by its positive correlation with the three scales purportedly tapping anxiety (GA, TA, LP). If one assumes nonoverlapping definitions of the characteristics measured, the obtained correlations would suggest the possibility that individuals who respond in a hostile manner in interpersonal situations are often reacting to anxiety with which they cannot adaptively cope. Also, of course, in many individuals, the expression of hostility leads to anxiety. The correlation of .56 between the H and LP scales suggest that insecurity and lack of protection in childhood may be an important factor in determining subsequent hostile behavior.

The other striking aspect of the correlation in Table 1 is that concerning the high negative *r*'s between scores on the D scale and all but one of the other scales. This finding lends further evidence to the hypothesis that the tendency to admit to undesirable personal characteristics (e.g., anxiety or hostility) is related to the degree of defensiveness exhibited by the individual. From a different point of emphasis, the anxious individual seems prone to admit to socially undesirable characteristics.

Verbal Conditioning

In order to test the hypotheses concerning the relationship between Autobiographical Survey scores and verbal conditioning, an analysis of covariance was performed for each scale of the survey, using the number of responses of the reinforced class (Verbal Activity) in Trials 1–20 as the control variable and the number of responses of the reinforced class during Trials 21–160 as the criterion variable. For the purpose of these analyses, the distribution of 60 scores for each scale was divided as nearly as possible into lower, middle, and upper thirds.

It had been predicted that *S*s with high scores on the GA, TA, and LP scales would give a significantly greater number of conditioned responses on Trials 21–160 than would *S*s with low scores on these scales. When the three LP groups were compared by analysis of covariance, an $F = 5.13$ ($p < .01$) was obtained. The adjusted means for these groups are presented in Table 2, and examination of them indicates that *S*s with low scores on the LP scale perform at a significantly lower level than do other *S*s in the LP score distribution. Virtually the identical finding was obtained for the TA scale. An $F = 3.49$ ($p < .05$) was obtained in this case; and as Table 2 indicates, the adjusted means for the middle and upper TA groups are significantly higher than for the lower TA group. The results with respect to the upper, middle, and lower GA groups failed to attain significance ($F = 1.68$, $p < .20$). However, the pattern of adjusted means was essentially the same as that obtained for the TA and LP scales. In general, then, the results with respect to these three scales provide support for the hypothesis that *S*s with relatively higher scores tend to adopt responses reinforced by *E* more readily than do low scores on three scales of the Survey.

The prediction had also been made that high scores on the D and H scales would be related to lowered performance levels in the verbal conditioning situation. This prediction was supported in the case of the D scale ($F = 4.65$, $p < .025$). The adjusted means (Table 2) for *S*s in the lower and middle thirds of the distribution of D scores were significantly higher than the mean for the high D group. It thus appears that degree of defensiveness is related to the extent to which *S*s adopt responses reinforced by *E*.

TABLE 2
MEANS ADJUSTED BY COVARIANCE ANALYSIS FOR HIGH, MIDDLE, AND LOW TA, LP, AND D GROUPS

	Adjusted Means		
	High	Middle	Low
TA scale	61.53	60.02	47.59
LP scale	64.90	59.93	47.08
D scale	48.09	59.51	64.14

The results with respect to the H scale, however, were not significant, and the prediction that high H scores would be related to poor verbal conditioning was not supported.

Psychotherapists' Ratings

An exploratory effort was made to determine if, as had been hypothesized, *S*s behave to some extent similarly in the psychotherapy as in the verbal conditioning situation. The method involved having therapists rate their patients on a number of variables, correlating these ratings with Autobiographical Survey scores, and also relating the therapist ratings to verbal conditioning scores.

To the extent that compliance with authority is a means of reducing tension and anxiety, it was predicted that therapists' ratings of the degree of compliance with which a patient accepted interpretations of the therapist would be positively correlated with the anxiety scales of the Survey. If this were so, it would also be expected that *S*s rated as being highly compliant would condition faster than would *S*s rated as being noncompliant. Biserial correlations were obtained between the ratings and scores on GA, TA, and LP for 38 of the 60 *S*s in this experiment. Significance was obtained only for the relationship involving LP ($r_b = .38$), although the GA and TA correlations were also positive and approached significance. Concerning the relation of verbal conditioning scores and therapists' ratings, an analysis of covariance for groups divided into high, middle, and low compliance ratings showed that the low compliance group performed at a significantly lower level in verbal conditioning than did the high and middle groups.

Psychotherapist ratings on two other variables (dependence and hostility as exhibited in psychotherapy sessions) were found to be unrelated to the Survey and also unrelated to performance in the verbal conditioning situation.

DISCUSSION

In reviewing the results of this experiment, cautious optimism would seem to be appropriate. The word optimism is used because the positive results of this study clearly demonstrate the relevance and importance of controlling individual personality differences in verbal conditioning experiments. The present results, together with those reported by Tatz (10), indicate that in learning situations involving reinforcement by *E*, *S*s bring into the experimental situation attitudes and anticipations that directly affect the extent to which they "accept" *E*'s reinforcement or use these reinforcements as a basis for modifying their own behavior.

Further, the majority of the predictions made with reference to the relationship of scores on the various scales of the Survey and scores on the conditioning series were borne out. These predicted relationships are generally assumed clinically and form a basis for such clinical predictions as of therapeutic success. Thus, the present results also lend support to the analogy drawn between *S*'s response to *E*'s reinforcement and a patient's response to a therapist's interpretations. Although the present findings are by no means conclusive, it does seem that a line of attack within the area of research in psychotherapy might be opened which could conceivably make good use of such an analogy. With the many methodological problems besetting the investigator interested in the psychotherapeutic process per se, it may be tactically prudent to concentrate a degree of energy on the study of nonclinical situations which possess characteristics similar to those obtaining in the clinic.

Several relevant questions are raised by the interrelationships found among the various scales of the Autobiographical Survey used in the study. 1. Are the highly related scores based on unclear or overlapping definitions of these personality characteristics? 2. Do these characteristics, if they are adequately defined, lack independence or are certain meaningful relationships present? 3. Might a combination of both 1 and 2 occur? 4. What characteristics differentiate individuals whose scores on the various Survey scales do not conform to the typical interrelationships shown by the groups of *S*s as a whole? For example, what are the characteristics of individuals whose scores on the various types of anxiety scales differ markedly, e.g., high scores on TA but low on GA or LP?

Caution is to be urged in generalizing on the basis of the present findings. It would seem of considerable importance to replicate the

present findings with clinical and nonclinical groups. No attempt was made in the present experiment to compare the verbal conditioning performance of groups of patients differing in diagnostic classification. Would, for example, the presumed greater social awareness of neurotics as compared with psychotics reflect itself in a greater degree of verbal conditioning? With respect to a comparison of the over-all performance of the present group of psychiatric patients with a nonpsychiatric population, the present results are comparable to those reported by Sarason (8), who devised the verbal conditioning procedure used in the present study. Studying VA nonpsychiatric patients, she obtained a considerably higher over-all level of conditioning than was obtained in the present study. Quite likely, the differences in patient population between her study and the present one is of great importance in bringing about this difference in level of performance. What precisely are the factors bringing about the differences in level of conditioning in psychiatric and medical patients is at this point unclear. Another possibly relevant factor accounting for this difference between the two studies is the sex difference between experimenters.

SUMMARY

1. The following measures were available for a group of psychiatric patients in active treatment: (*a*) Scores on the six scales (General Anxiety, Test Anxiety, Lack of Protection, Defensiveness, Need for Achievement, and Hostility) of a personality inventory, the Autobiographical Survey. (*b*) Ratings by the patients' psychotherapists of the degree of compliance, dependence, and hostility displayed by the patients in therapy sessions. (*c*) Level of performance attained in a verbal conditioning task in which *E* systematically reinforced all responses made by *S* which fell in the category of Verbal Activity (i.e., verbs indicating use of the vocal apparatus).

2. The six scales of the Autobiographical Survey were found to possess reasonable reliabilities. Intercorrelations among the six scales indicated positive relationships among all the scales except the Defensiveness scale, which correlated negatively wth the other scales.

3. High scores on the Test Anxiety and Lack of Protection scales were found to be significantly associated with higher levels of verbal conditioning. On the other hand, high Defensiveness scores were found to be related to relatively poor verbal conditioning.

4. Patients rated by their psychotherapists as being very compliant were found to perform in the verbal conditioning situation at a significantly higher level than patients rated as being relatively noncompliant.

5. The results were discussed in terms of previously made predictions and in terms of the feasibility of studying factors affecting behavior in psychotherapy in simpler, more easily controllable situations.

REFERENCES

1. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.
2. FREUD, S. *The problem of anxiety*. New York: W. W. Norton, 1936.
3. GORDON, E. M., & SARASON, S. B. The relationship between "test anxiety" and "other anxieties." *J. Pers.*, 1955, **23**, 317–323.
4. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
5. HATHAWAY, S. R., & MCKINLEY, J. C. *Minnesota Multiphasic Personality Inventory*. New York: Psychological Corp., 1943.
6. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
7. MARK, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 185–189.
8. SARASON, B. R. The effects of verbally conditioned response classes on post-conditioning tasks. Unpublished doctoral thesis, Indiana Univer., 1956.
9. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.
10. TATZ, S. Symbolic mediation in "learning without awareness." Unpublished doctoral thesis, Yale Univer., 1956.
11. TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
12. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1944.
13. VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.
14. VERPLANCK, W. S. The operant, from rat to man: An introduction to some recent experiments on human behavior. *Trans. N. Y. Acad. Sci.*, 1955, **17**, 594–601.

Received May 31, 1957.

INTERRELATIONS AMONG MEASURES OF PERSONALITY ADJUSTMENT IN NONCLINICAL POPULATIONS[1]

FRED E. FIEDLER, JOAN S. DODGE,[2] ROBERT E. JONES,[3] AND EDWIN B. HUTCHINS

University of Illinois

A LARGE number of studies, reported in the literature, require the evaluation of psychological adjustment or of changes in adjustment, whether the study deals with consequences of situational factors, or of psychotherapeutic or psychosurgical intervention.

Most theoretical and empirical work in this area treats the person's adjustment as falling on a unidimensional scale. Thus, Sappenfield writes, "Adjustment involves a continuous variable, and therefore, the evaluation of individuals in terms of this variable should not be limited to the two extremes, [of adjusted and maladjusted]" (19, p. 24). A similar point of view is expressed or implied by other writers in this area (e.g., 2, p. 15; 6, p. 554; 23, p. 102).[4] It has gained considerable theoretical support from the writings of Goldstein (10) and Rogers (17, p. 510) who see the individual reacting to his environment as a total organism, and who see poor adjustment in one area of life as necessarily affecting all other areas of personality. Improvement or deterioration of adjustment should, therefore, take place on all fronts.

Whether adjustment is in fact unidimensional in character, or whether a person may achieve adjustment in one area of his life irrespective, or at the expense of, his adjustment in another is a question for research to which the present report hopes to contribute. It asks how various adjustment indices intercorrelate with one another in samples of subjects (*Ss*) whose adjustments, as measured by a variety of indices, range from excellent to relatively poor.

Other workers in this area have previously raised this question. Ellis (5) has shown that objective personality questionnaires are of dubious value in distinguishing between adjusted and maladjusted individuals, and that intercorrelations among similar personality questionnaires are generally low or insignificant. Tindall (21), in a more recent paper, likewise reports low intercorrelations among 16 adjustment indices. On the other hand, Tucker (22) reports positive correlations among change in adjustment measures, and Cartwright and Roth (1), working with post-therapy ratings and indices, have found that 18 of 45 correlations among adjustment scores were greater than .46 (beyond the .01 level) although pretherapy scores do not intercorrelate. To summarize, the status of our knowledge concerning indices of adjustment is by no means clear. The present report deals with a fairly wide variety of frequently used measures of adjustment and change in adjustment, and it compares results from four studies which used subjects (*Ss*) who were unselected in terms of clinical symptomatology.

The data were gathered in connection with a research program on "quasi-therapeutic" relations in small groups. The studies were conducted on the assumption that any interpersonal relationship might be helpful or detrimental to adjustment, and that the adjustive relations could be differentiated from those which would lead to deterioration of adjustment.

The specific design of the various studies will be considered in a forthcoming report. The present paper limits itself to a brief description of the samples, the methods, and the research setting in each of the four studies; and the presentation and interpretation of data relevant to our inquiry.

METHOD

The Four Studies

We studied four different groups, two of which consisted of college students, and two of military personnel, viz., army tank crews and anti-aircraft artillery (AAA) units. Data were collected twice in each of the studies, so that we could obtain scores on adjustment indices at

[1] This study was conducted by the Group Effectiveness Research Laboratory, Department of Psychology, University of Illinois, under Contract DA-49-007-MD-569 with the Office of the Surgeon General. Reproduction in whole or in part is permitted for any purposes of the United States government.

[2] Now at the Graduate School of Business and Public Administration, Cornell University.

[3] Now at the VA Mental Hygiene Clinic, Huntington, W. Va.

[4] This is, of course, also the underlying assumption wherever a number of adjustment scores are averaged or combined.

the first and second testing sessions, as well as scores indicating changes in adjustment.

The Dormitory study. The first study was conducted on a group of 107 male college students living in a new university dormitory.[5] Most *Ss* had entered college at the beginning of the semester in which the data were collected. About half of the students were freshmen, the remainder being transfer and incoming graduate students. *Ss* were assigned to their rooms in order of their application for housing. All *Ss* were tested during the fifth week of the semester, and again during the eleventh week. Our analyses are based on 87 *Ss* for whom complete data were available.

The Barracks study. Three years after the Dormitory study, similar data were collected on a sample of students living in 13 temporary barracks. Each barracks, which is divided into four-man suites, holds a maximum of 16 men. Although a total of 172 *Ss* were tested, analyses were restricted to the 102 men for whom complete data could be obtained at both testing sessions. These *Ss* were tested during the third and twelfth weeks of the semester.

The Tank Crew study. This investigation was conducted on members of a tank battalion during the crew training phase. *Ss* were enlisted men with an average age of 22 years. Of the 160 men initially tested, 49 were married, three divorced. Average education was 9.7 years, mean Army General Classification Test score was 90. All men were new to each other and to the particular type of tank training. This group was thus considerably more heterogeneous than the student groups in the previously mentioned studies. Upon retest, the sample size was reduced from 160 to 95 due to transfers, discharges, and other causes.

The Anti-aircraft Artillery study. The last group of *Ss* came from operational AAA units. Here again there was considerable heterogeneity among *Ss*. Average age of the men, all of whom were enlisted, was 23 years, average education 11.2 years, mean Army General Classification Test score was 101. Of the 200 men in our sample, 36 were married.

The AAA crews were "old" units, that is, these men were no longer in training, but were performing their regular duties within the Air Defense Command. As with tank crews, the loss of data, due to furloughs, sick call, or turnover in our particular sample was considerable. In the 12-week period between testings, the original sample shrank from 485 men to 200 for whom complete data were available. Our analyses are based on this group of 200 *Ss*.

Criteria of Adjustment

As a rule, the development of adequate criteria of adjustment presents a very difficult problem. One major source of this difficulty arises from the fact that each theoretical orientation has defined adjustment or integration on the basis of different assumptions and by different operations. However, it is perhaps somewhat easier to develop adequate adjustment measures in a group setting than when the individual is considered by himself, as is the case in most studies on individual therapy. The individual in a group situation is almost continuously interacting with others, and he is thus under almost constant observation by his fellow group members. In addition, each man shares the environmental stresses of others in the group, and his reactions to these stresses are therefore comparable to those of other group members.

We shall here deal with three classes of criteria. These are (*a*) subjective indices of adjustment which depend on *S*'s self-reports, (*b*) group acceptance criteria which are based on group members' evaluations or descriptions of *S*. and (*c*) personal effectiveness criteria indicative of the individual's ability to operate effectively. Most of these indices, or similar ones, have been used by other investigators in this area.

Subjective indices of adjustment. These criterion scores were derived from *S*'s self descriptions. In particular we employed scale sheets on which *S* described himself as he thought he really was, and how he would ideally like to be (his ideal self), and he described various other members of his unit. These descriptions were made on a series of identical scale sheets, each of which contained the same 20 items.[6] Items were six-point scales bounded by pairs of polar adjectives. For example:

Friendly ___:___:___:___:___:___ Unfriendly
Stable ___:___:___:___:___:___ Unstable

In the Dormitory and Tank Crew studies, *Ss* described all other men in their units; in the Barracks and AAA studies, where each unit was considerably larger, *Ss* were asked to describe only a limited number of men. As we shall discuss below, these descriptions of others were used as indices of group acceptance.

Self-esteem scores could be obtained from self-descriptions by determining the average favorability of *S*'s self rating. As will be noted, each of the 20 items can easily be polarized: it is better to be friendly than unfriendly, stable than unstable, and so forth. A person who consistently describes himself in a more favorable manner is presumed to have higher *self-esteem*. Obviously, as is the case with all self reports and inventories, this score may reflect very superficial attitudes, such as a facade, or it may reflect a genuine degree of self-esteem on the part of the *S*.

Using the measure D (4, 15) we obtained a second index, called *Self Satisfaction*, which is defined as the degree to which *S*'s self-description is similar to his description of how he ideally would like to be. This index has been utilized by one of the authors (Fiedler, 7) in studies of patient-therapist interactions. It has been extensively used by Rogers and his co-workers to demonstrate movement in client-centered therapy (1, 16).

In addition, an abbreviated form of the Taylor Manifest Anxiety Scale (20), which has become one of the most popular inventories in recent studies of adjustment (9, 12), was given to all *Ss* save those in the Dormitory sample. In the AAA study, the *General Army Adjustment* scale, described by Havron, Fay, and Goodacre (11) was also administered.

Group acceptance criteria. Two measures were obtained. One, *Esteem by Others*, is derived from rating

[5] We are indebted to Melvin Manis for permission to utilize his data for purposes of our project. His own study is reported in (14).

[6] The Dormitory study used seven-point, 24-item scales.

scales on which *S*s were asked to describe others in their group. It is analogous to the Self-Esteem score, and again is the average item score with which a particular *S* is described. The second index, *Sociometric Status*, was obtained by asking *S*s to name those in their group whom they preferred most, in whom they would confide personal problems, whom they would choose as a leader, etc. A pooling of these highly intercorrelated items yielded a sociometric preference index. The more an *S* was chosen, the higher is his status in the group. (Cf. French (8), Pepinsky, Siegel, and Van Atta (16), and Cox (3) whose studies have related sociometric status to some aspects of adjustment.)

Personal effectiveness criteria. Three personal effectiveness criteria were available for the college student samples. The first of these is based on the frequency of the individual's *Health Center* visits. It assumes that the anxious, maladjusted person will experience relatively more symptoms of a psycho-physiological nature than will one who is adjusting. He will also be more aware of his aches and pains, partly because of hypochondriacal defense tendencies, partly because general anxieties will tend to augment pain through increased tension.

Student Counseling Bureau visits for personal maladjustments were similarly tabulated for our *S*s as an indication that the individual feels that he cannot solve his problems without professional help.

Finally, we collected students' *Grade Point Averages* (GPA) for the semester of our study. In the Barracks study, we were also able to obtain mid-semester grades which are given to freshmen on an advisory basis. Collection of these data permitted us to compute change in grade point average for this sample.

In the AAA units, adequate individual proficiency scores were not available. Sick-call data were obtained but, for a variety of reasons, such as unequal accessibility of the dispensary to various units, they do not appear to be adequate for research purposes. Ratings on disciplinary offenses appeared to be usable, however. These data were obtained in the AAA study and provide some indication of the individual's ability to adjust to the Army.

Measures of Change in Adjustment

As indicated above, all *S*s were tested twice, with intervals between testing sessions ranging from six weeks to 13 weeks in the various samples. Since any change score necessitates two measures on the same individual, these scores are available only for subjective indices, group acceptance criteria, and for grade point average for the Barracks study.

Reliability of Scores

The relation between two indices of adjustment cannot be interpreted adequately unless we know the reliability of each of the scores. Table 1 lists the corrected split-half reliabilities. Only one estimate is here presented. Where more than one reliability was computed, the estimate based on the larger N is shown. Relatively little variation was found from sample to sample. Since some scores, such as Health Center Visits or Counseling Bureau Visits, do not lend themselves to this type of statistical analysis, these indices must be interpreted with considerably more caution.

TABLE 1

SPLIT-HALF RELIABILITY COEFFICIENTS[a] OF PERSONALITY ADJUSTMENT INDICES

Index	Sample	N	r
Self-satisfaction	Barracks	169	.76
Self-esteem	AAA	397	.83
Mean Esteem by Others	Dorms	86	.70
Sociometric status	Dorms	99	.94
Taylor Anxiety	AAA	200	.91
General Army Adjustment	AAA	200	.73

[a] Coefficients are the corrected r's based on the Spearman Brown Prophecy formula.

CORRELATIONS AMONG CRITERIA OF LEVEL OF ADJUSTMENT

If adjustment is to be treated as a unitary trait or general attribute, it is expected that specific measures of its various operational definitions will be correlated. An individual showing superior adjustment on one index should be above average on others as well. A lack of such co-variance would suggest that various adjustment indices measure very specific components of adjustment, or else, that adjustment in one area of behavior does not necessarily imply adjustment in a different area.

Intercorrelations of Adjustment Indices at Time II

Table 2 presents the intercorrelations among the various adjustment indices at the time of the second testing. As can be seen, all correlations are very low, even when highly reliable internal criteria, such as Taylor Anxiety scores and self-satisfaction or self-esteem scores, are involved.

It must be recognized, of course, that some maladjustment symptoms are substitutive rather than supplemental or additive. A person who goes to the Health Center for psychosomatic symptoms is less likely to consult the Counseling Bureau as well. Yet, we would expect that someone who seeks help at the Counseling Bureau or the Health Center or who performs poorly in school would be dissatisfied with himself. As can be seen, however, this latter expectation is not borne out by our data.

Although relationships between indices presented here are not artifactual in the mathematical sense, all indices based on self-description are psychologically related. A person who

TABLE 2

SUMMARY OF RELATIONSHIPS BETWEEN INDICES OF ADJUSTMENT OBTAINED AT THE SECOND TESTING

Indices Related	Dorms $N = 87$	Barracks[d] $N = 102$ [$N = 71$ (Est. by O.)]	Tanks $N = 95$ [$N = 52$ (Est. by O.)]	AAA $N = 200$
Self-Satisfaction and:				
Self Esteem	.60[a]**	.55[a]**	.36[a]**	.44[a]**
Sociometric Status	.09	.09	.00	.10
Mean Esteem by Others	−.02	.15	−.14	−.09
Taylor Anxiety	—	.34[b]*	.37[b]**	.30[b]**
Health Center Visits	.00	.19	.10	—
Grade Point Average	.17	.30	—	—
Counseling Bureau Visits	.15	.16	—	—
Annoyance Scale	—	—	.08	—
Disciplinary Ratings	—	—	—	.01
General Army Adjustment	—	—	—	.15*
Self Esteem and:				
Sociometric Status	.20	.10	.04	.09
Mean Esteem by Others	.22*	09	−.20	.02
Taylor Anxiety	—	.25[b]	.43[b]**	.34[b]**
Health Center Visits	.22*	−.06	−.09	—
Grade Point Average	−.14	.28	—	—
Counseling Bureau Visits	.08	.22	—	—
Annoyance Scale	—	—	−.10	—
Disciplinary Ratings	—	—	—	.11
General Army Adjustment	—	—	—	.16*
Sociometric Status and:				
Mean Esteem by Others	.38[c]**	.42[c]*	.34[c]*	−.17[c]**
Taylor Anxiety	—	−.09	.09	.04
Health Center Visits	.06	−.04	.02	—
Grade Point Average	.05	.31*	—	—
Counseling Bureau Visits	.10	.00	—	—
Annoyance Scale	—	—	.01	—
Disciplinary Ratings	—	—	—	.33**
General Army Adjustment	—	—	—	−.03
Mean Esteem by Others and:				
Taylor Anxiety	—	−.02	−.13	−.07
Health Center Visits	.15	.08	−.09	—
Grade Point Average	.03	.14	—	—
Counseling Bureau Visits	.23*	.06	—	—
Annoyance Scale	—	—	−.07	—
Disciplinary Ratings	—	—	—	.24**
General Army Adjustment	—	—	—	.12
Taylor Anxiety and:				
Health Center Visits	—	.16	.07	—
Grade Point Average	—	.18	—	—
Counseling Bureau Visits	—	.05	—	—
Annoyance Scale	—	—	−.01	—
Disciplinary Ratings	—	—	—	.04
General Army Adjustment	—	—	—	.27**
Health Center Visits and:				
Grade Point Average	−.10	−.06	—	—
Counseling Bureau Visits	.06	.02	—	—
Annoyance Scale	—	—	−.07	—
Grade Point Average and:				
Counseling Bureau Visits	.03	.27	—	—
Disciplinary Ratings and:				
General Army Adjustment	—	—	—	.11

[a] Spuriously high relation based on self-descriptions.
[b] Both variables are self-descriptions.
[c] Both variables based on descriptions by the same others.
[d] r's estimated from ϕ coefficients.
* Significant at the .05 level.
** Significant at the .01 level.

describes himself as well adjusted on the Taylor Manifest Anxiety Scale is likely to describe himself also as favorable on the self-description sheet. Hence, his Self-Esteem score, and his Self-Satisfaction, will also be relatively high. Under these conditions, one would actually expect higher correlations than occurred here. Similarly, Sociometric Status and Esteem by Others are psychologically related both being ratings given to *S* by his fellow group members. Since they are based on responses from the same raters, it is again noteworthy that the relations are so low. Inspection of scatter diagrams reveals no evidence of curvilinearity in our data.

Intercorrelations were also computed for indices obtained at the first testing session. These are essentially similar to those found in Table 2. (A comparison is presented in Table 4 which summarizes the median intercorrelations obtained at the first and second testing session and for change scores which are presented below.)

As can be seen from Table 2, there is relatively little similarity in the pattern of intercorrelations from study to study. These data,

TABLE 3

SUMMARY OF RELATIONSHIPS BETWEEN INDICES OF ADJUSTMENT (CHANGE SCORES)

Indices Related	Dorms $N = 87$	Barracks[d] $N = 66$	Tanks $N = 95$ [$N = 52$ (Est. by O.)]	AAA $N = 200$
Self Satisfaction and:				
Self Esteem	.46[a]**	.43[a]**	.18[a]	.38[a]**
Sociometric Status	−.11	−.10	.00	.08
Mean Esteem by Others	.15	.03	.09	.04
Taylor Anxiety	—	.26[b]*	.24[b]*	.08[b]
Health Center Visits	—	—	.18	—
Annoyance Scale	—	—	.10	—
General Army Adjustment	—	—	—	.01
Grade Point Average	—	−.12	—	—
Self Esteem and:				
Sociometric Status	.08	−.14	.01	.04
Mean Esteem by Others	.00	.23	−.13	−.02
Taylor Anxiety	—	.23[b]	.01[b]	.09[b]
Health Center Visits	—	—	−.09	—
Annoyance Scale	—	—	−.10	—
General Army Adjustment	—	—	—	.00
Grade Point Average	—	−.08	—	—
Sociometric Status and:				
Mean Esteem by Others	.24**	.09[c]	.16[c]	−.09[c]
Taylor Anxiety	—	.26*	.05	−.03
Health Center Visits	—	—	−.15	—
Annoyance Scale	—	—	.10	—
General Army Adjustment	—	—	—	.00
Grade Point Average	—	.19	—	—
Mean Esteem by Others and:				
Taylor Anxiety	—	.04	.11	−.03
Health Center Visits	—	—	−.11	—
Annoyance Scale	—	—	.04	—
General Army Adjustment	—	—	—	.01
Grade Point Average	—	.01	—	—
Taylor Anxiety and:				
Health Center Visits	—	—	.13	—
Annoyance Scale	—	—	.13	—
General Army Adjustment	—	—	—	.25**
Grade Point Average	—	.27*	—	—
Health Center Visits and:				
Annoyance Scale	—	—	.01	—

[a] Spuriously high relation based on self descriptions.
[b] Both variables are self descriptions.
[c] Both variables based on descriptions by the same others.
[d] *r*'s estimated from ϕ coefficients.
* Significant at the .05 level.
** Significant at the .01 level.

TABLE 4
SUMMARY OF MEDIAN CORRELATIONS[a] OBTAINED FROM THE FIRST AND SECOND TESTING SESSIONS AND FROM CHANGE SCORES

	Dorms	Bar-racks	Tanks	AAA
First testing session	.16	−.05	.07	.11
Second testing session	.10	.14	.01	.11
Change scores	.08	.07	.05	.01

[a] In determining the median correlations r's which were spuriously high were not included.

as well as the generally low intercorrelations, indicate that factor analyses would not contribute materially to their interpretation.

Interrelations among Change in Adjustment Scores

Although we have not found relationships among level of adjustment scores, it is nevertheless possible that changes in adjustment over various aspects of personality might be related. Improvement on one criterion score could be accompanied by improvement on others irrespective of the level at which *S*s are observed to be at any one time. To determine whether or not co-variations in change scores do exist, all available change scores were intercorrelated. Table 3 presents these findings.[7]

If we exclude the relations which are artifactual in nature, the correlations are again uniformly low. In our samples, then, there is no evidence that improvement (or deterioration) on one criterion is accompanied by significant changes in other types of criterion scores. Nor are the relations very similar if we compare the profile or correlations in one study with that in another.

Obviously, the criteria of change do not vary together. A person who improves in his subjective feelings may or may not act in a more adjusted manner as judged by others, and he may or may not perform more adequately or efficiently in performing his job. A few individuals undoubtedly do change in all aspects of adjustment, but the possible identification of these individuals is a question for further research.

[7] It should be pointed out that change scores are subject to "regression to the mean" effects which can seriously distort our estimates of the "true" gain scores. Lord (13) has pointed out that these effects will be considerable when the standard deviations differ greatly at Time I and at Time II. Since this is not the case in our studies, it does not affect materially the correlations presented in this paper.

DISCUSSION

While there is, of course, no logical reason to expect correlation between measures such as sociometric status and self satisfaction, both are frequently considered indices of adjustment. Table 4 summarizes the median intercorrelations among the criterion scores in these studies. Less than half are above .10 and none exceed .16. Text books do not generally discuss the possibility that Adjustment A may be totally different from Adjustment B. Although Rogers and Dymond (18, p. 431–432) explicitly recognize this possibility (which contrasts with Rogers' earlier view), we cannot readily find this viewpoint in standard psychiatric texts, or in many of the texts on personality theory or abnormal psychology.

It is especially noteworthy that Cartwright and Roth (1) have reported fairly high posttherapy intercorrelations in clinically selected *S*s but low relations in pretherapy scores. This observation may point to the possibility that the structure of posttherapy adjustment in individuals who complete counseling or therapy may differ from that of *S*s who never felt the need for treatment. It could be possible that the individuals in therapy experience a more basic kind of anxiety, interfering more comprehensively with life adjustment problems. It is also possible that therapy typically aims for, and achieves, a rebalancing of defenses so that the individual is similarly adjusted in all areas of living. In any case, these discrepancies need to be further examined.

The generalizability of these findings may be challenged on the grounds that the sample of criteria chosen was not adequately representative of all adjustment areas. This possible criticism is, of course, a matter of definition. We recognize that we did not include many excellent adjustment inventories or indices which have been utilized in the past. However, the fact that Tindall's (21) and Ellis' (5) findings agree with ours adds to the weight of our evidence, as does the fact that our findings are very similar in four different samples of *S*s.

A more important question is whether the time span between the first and the second testing is sufficiently long enough to permit us to see co-variations in change scores. We recog-

nize this to be a definite limitation of our data, although studies not yet reported have shown that some "quasi-therapeutic" changes do take place within this time.

SUMMARY

This paper has presented intercorrelations among a variety of indices which have been used as measures of personality adjustment. The most important fact that emerges is the general lack of correlation among different indices—even among those which are reliably measurable and which could be expected to correlate with each other. Thus, our data yield no evidence justifying the assumption that adjustment, in its present state of definition, should be considered a unitary trait in clinically unselected populations.

In light of our data, we must recommend that subsequent research be confined to specific measures when inferences are to be drawn about personality adjustment or personality changes. It should also be recognized that results from clinically selected, or post-therapy groups, may be different from those obtained on *S*s who are unselected from the point of view of presenting symptoms, or the fact that they expressed a need for psychotherapeutic help.

REFERENCES

1. CARTWRIGHT, D. S., & ROTH, I. Success and satisfaction in psychotherapy. *J. clin. Psychol.*, 1957, **13**, 20–26.
2. COLEMAN, J. C. *Abnormal psychology and modern life.* (2nd ed.) New York: Scott Foresman, 1956.
3. COX, F. N. Sociometric status and individual adjustment before and after play therapy. *J. abnorm. soc. Psychol.*, 1953, **48**, 354–356.
4. CRONBACH, L. J., & GLESER, GOLDINE C. Similarity between persons and related problems of profile analysis. *Psychol. Bull.*, 1953, **50**, 456–473.
5. ELLIS, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, **42**, 385–440.
6. FENICHEL, O. *The psychoanalytic theory of neuroses.* New York: Norton, 1945.
7. FIEDLER, F. E., & SENIOR, KATE. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. *J. abnorm. soc. Psychol.*, 1952, **47**, 790–796.
8. FRENCH, R. L. Sociometric status and individual adjustment among naval recruits. In D. Cartwright & A. Zander (Eds.), *Group dynamics.* Evanston, Ill.: Row Peterson, 1954.
9. GALLAGHER, J. J. Manifest anxiety changes concomitant with client centered therapy. *J. consult. Psychol.*, 1953, **17**, 443–446.
10. GOLDSTEIN, K. *Human nature in the light of psychopathology.* Cambridge: Harvard Univer. Press, 1940.
11. HAVRON, M. D., FAY, R. J., & GOODACRE, D. M., 3rd. *Research on the effectiveness of small military units.* PRS Report No. 885. Institute for Research on Human Relations, 1951.
12. HOLTZMAN, W. H., CALVIN, A. D., & BITTERMAN, M. E. New evidence for the validity of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853–854.
13. LORD, F. M. The measurement of growth. *Educ. psychol. Measmt*, 1956, **16**, 421–437.
14. MANIS, M. Social interaction and the self concept. *J. abnorm. soc. Psychol.*, 1955, **51**, 362–370.
15. OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.
16. PEPINSKY, H. B., SIEGEL, L., & VAN ATTA, E. L. The criterion in counseling: A group participation scale. *J. abnorm. soc. Psychol.*, 1952, **47**, (Suppl.) 415–419.
17. ROGERS, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
18. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.
19. SAPPENFIELD, B. *Personality dynamics.* New York: Knopf, 1956.
20. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
21. TINDALL, R. H. Relationships among indices of adjustment status. *Educ. psychol. Measmt*, 1955, **15**, 152–162.
22. TUCKER, J. E. Measuring progress in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* Pennsylvania State Coll., 1953.
23. WHITE, R. W. *The abnormal personality.* New York: Ronald Press, 1956.

Received May 31, 1957.

INTERACTING EFFECTS OF CONTROL, MOTIVATION, GROUP PRACTICE, AND PROBLEM DIFFICULTY ON ATTEMPTED LEADERSHIP[1]

BERNARD M. BASS, MARGARET W. PRYER, EUGENE L. GAIER, AND AUSTIN W. FLINT

Louisiana State University

THE differentiation of *attempted* from successful and effective leadership was suggested by Hemphill[2] and applied in the construction of two distinct theories of leadership, one by Hemphill and one by Bass (2). For the latter, leadership is attempted whenever one member of a group behaves in a manner directed toward changing any behavior of another. Hemphill's definition is more restricted. Leadership attempts refer only to acts by one member aimed at restructuring the existing pattern of interaction between that member and others.

The broader conceptualization permits the acceptance of almost any response in initially leaderless discussions as attempted leadership. Participation in leaderless discussions is equated with attempted leadership, since it is assumed that almost all talk in initially leaderless discussions is directed toward influencing other members primarily by giving opinions, suggestions, information, and by summarizing, clarifying as well as organizing, the topic discussed (1).

One member, A, *controls* another, B, if he can give or get for B what B wants. If B's wants are strong, A has *power* to change B's behavior by giving or denying B what B wants. Therefore, A can successfully lead B, if leadership behavior is defined as acts directed toward changing another's behavior. But A will be successful only if B's motivation to obtain what A can give is fairly high.

If A expects to be successful, he is more likely to attempt leadership. Hence, under conditions of high motivation, A's control results in power to lead B and A will attempt to do so. However, if motivation of A and B are low, A's control will not yield power and A will not attempt more leadership.

A second cause-effect mechanism that relates control and motivation produces the same outcome: Member A may be expected to use his high control only if he is highly motivated. Either rationale supports the deduction that under conditions of strong motivation, high control members will attempt more leadership than low control members. Under conditions of weak motivation, differences should become negligible or completely disappear.

The major purpose of the present study was to test this hypothesis that control yields attempted leadership but only under conditions of high motivation. A second purpose was to examine the effects of group practice on the relations of control, power, and attempted leadership. Continued experience was expected to reinforce the attempts of high power members to lead.

A third purpose was to examine the effects of problem difficulty on attempted leadership. Easy problems require relatively little leadership, and little should be attempted; but attempted leadership should be more frequent, especially by the more powerful members, as problems become more difficult (but not so difficult that the group withdraws from trying to solve them).

METHOD

Manipulating Motivation by Selection

On an a priori basis, it was believed motivation could be varied most successfully by classification of *S*s according to their varying motivation to attain some important life goal. At the end of the sophomore year, the land-grant college student has fulfilled the requirements of the federal government for participation in ROTC training. However, if the student chooses, and is found qualified, he may remain in the ROTC program two additional years, and receive an officer's commission at the end of this period. Students vary widely in their desire to enter Advanced ROTC. We obtained varying levels of motivation for test performance by subjecting students differing in stated desire to enter Advanced ROTC, to an ROTC "screening test." To sample ade-

[1] This study was supported by Contract N7 ONR 35609, Group Psychology Branch, Office of Naval Research.

[2] Unpublished manuscript by J. K. Hemphill, "A proposed theory of leadership in small groups."

quately students with low motivation, all students, regardless of motivation, were ordered to take the test. *Ss* were instructed that the "teamwork test" might aid in determining whether or not the student was qualified for acceptance for advanced training. It was expected that students strongly interested in entering Advanced ROTC would be more highly motivated to perform well in the test situation than students who lacked such interest.

About two months before the present study was initiated, the entire sophomore class was given a questionnaire composed of seven five-step rating scales concerning patriotism and military service. Six were "dummy" scales. The key item was: "How important is it for you to enter Advanced ROTC?" The 342 sophomores who responded, did so in the following percentages:

Response	Percentage Responding
A. A great deal	21%
B. Fairly much	18
C. To some extent	34
D. Comparatively little	11
E. Not at all	16

Several analyses support the inference that response to the key question provided a valid assessment of motivation:

Attitude toward military career. Response to the key question was consistent at the .01 level of confidence with response to the item "Right now, how desirable is a career in the military service to you?"

Application for advanced ROTC. The following percentage of each category or response to the key question actually applied for Advanced ROTC during the next three months following completion of the questionnaire:

	Response to Key Question				
	A	B	C	D	E
Percentage applied	80	71	43	17	0
Percentage did not apply	20	29	57	83	100

For experimental use, the number of motivation levels was reduced to three approximately equal samples as follows:

Motivation Level	Responses	Percentage Responding
High	A and B	39%
Medium	C	34
Low	D and E	27

Failure to report. The lower two levels of motivation were less likely to appear for the "screening" examination. In fact, in order to reduce to a minimum the loss of groups, because someone did not appear for testing, it was necessary to schedule two alternate *Ss* for the Medium (C) and Low (D and E) motivation level groups, while only one alternate was necessary for the high level (A and B) groups. Despite this precaution, the following percentages of scheduled groups were not tested because of failure to appear of at least five men *including* the alternates.

Motivation Level	Responses	Percentage of Scheduled Groups Lacking Five Men to Be Tested
High	A and B	13.3%
Medium	C	20.0
Low	D and E	46.7

Fortunately, many of the men from the aborted groups could later be rescheduled in different groups.

Reported effort. About a week following the last testing session, a questionnaire was sent to all participating *Ss*. Of these, 198 returned the questionnaire. Motivation level was significantly related to effort according to self-ratings to the question "How hard would you say you tried to do as well as possible on the teamwork test?" Thus, 59% of the "highs" reported they had tried a great deal, 42% of the medium sample so reported and only 27% of the "lows" indicated such effort.

Griping behavior. Following a suggestion drawn from Willerman's (6) work on involvement in groups that low motivation *Ss* would place emphasis on taking the test, while *Ss* with high motivation would emphasize high score barriers, *Ss* were asked on the posttest questionnaire to comment on any aspect of the testing procedure. A content analysis found that the "highs" were more likely (at the .01 level of confidence) to complain about difficulties in earning a high score while "lows" griped about having to take the test. "Lows" were also less likely to return the questionnaire.

Manipulating Control

Amount of control of individual members as well as their variance in control were manipulated by experimental treatment of their status.

Amount of control. Before actual testing, each of the five members of the group drew from a deck, a card with one to four "Xs" marked on it. The number of "Xs" a member drew signified how much control he had. All were informed that the number of "Xs" determined

how much weight would be carried by each *S*'s evaluation of the other members' performance at the end of the screening examination. For example, a member who drew four "Xs" would carry four times as much weight in determining everyone else's final ratings on adequacy of test performance, as compared with a person who drew one "X."

Variance control. The extent to which members varied in control within a given group was regarded as a significant determiner of how members would react. The greater the differentiation among members, the more distinct should be differences in behavior due to control. To manipulate *control variance,* the decks of cards were "stacked" in one of three ways. In one-third of the groups tested, the five members drew from a 20-card deck where every card was marked with a single "X." In one-third of the groups, the five members drew from a five-card deck which contained three cards with one "X," one card with two "Xs" and one card with three "Xs." In the remaining one-third of the groups, the five members drew from a five-card deck which contained four cards with single "Xs" and one card with four "Xs." The three types of groups had variances in control of 0.00, 0.64, and 1.44 respectively.

Measuring Attempted Leadership

Attempted leadership was defined as the amount of time spent talking as measured by a "throat microphone-to-electric clock" circuit. As soon as the *S*s were seated for testing, each *S* fastened around his neck a throat microphone, which was connected to an electric clock circuit. The time in seconds for each of ten brief problem-solving discussions was recorded by an assistant.

The Group Task

Each group was administered ten problems. In each problem, the five members first privately ordered their opinions of the correct rank order of the familiarity of five words. They then carried on a group discussion to reach a group decision about the ranks. Again, they privately ranked the words. The correct rank order according to a word familiarity list constructed by Noble (5) was then reported to the participants (3).

Problem difficulty. An easy problem was constructed by assembling from Noble's list words that widely differed in familiarity and had a high mean familiarity as a set. A difficult problem was constructed using unfamiliar words differing little from each other in their familiarity. The problems were pretested on 100 Ohio State University freshmen. On the basis of the average correlation between each of these *S*'s initial private rank orderings of the five words of each problem, pairs of problems at each of five levels of difficulty were identified.

Randomized latin squares were used to provide each group with a random sequence of problems of varying difficulty. One additional restriction was imposed. If a problem appeared in the first half of the ten problems presented to a group, its equally difficult mate appeared somewhere in the second half. This provision was to insure that the overall difficulty of the second half of the test would be comparable to the difficulty of the first half.

Results

Original plans called for assessing 90 groups, 30 at each of the three levels of motivation. Practical considerations permitted the testing of only 60 groups, and in order to maintain proportionality to permit an unbiased two-factor analysis of variance of the effects of control and motivation on attempted leadership, we withdrew nine groups randomly (within the restrictions of the design) leaving 51 in the final analysis, distributed as shown in Table 1.

The data analysis will be presented in two parts. First, intergroup differences in attempted leadership (or total amount of participation) under the varying conditions of motivation and control variance will be tested. Then we will consider how an individual differed from his own group mean in attempted leadership as a consequence of his amount of control over the others in his own group. In the first case we will examine the absolute differences between groups of different control variance ignoring relative variations within the groups. In the second type of analysis, we will examine the relative differences between members of different control, ignoring the absolute differences between groups. In both cases, we will examine trends in obtained relations as a function of group practice and problem difficulty.

Analysis of Variance of Groups

Table 2 displays the average member's attempted leadership or amount of participation in the 51 groups under one of the three levels of motivation and three levels of variance in control averaged for all ten problems. According to Type III (4) analyses of variance, attempted leadership varied significantly at the .01 level with motivation. The "lows" in motivation talked only 58 per cent as much as the "highs," the "medium motivated" talked only 74 per cent as much as the "highs." The differences due to variance in control, to problem difficulty and to group experience were accounted for by the fluctuations between groups or within groups on repeated trials.

We expected that decisions would be reached most quickly with least total participation under high variance in control, but the trend —although not significant—was in the reverse direction. High variance in control, particularly under high motivation, yielded the greatest total amount of attempted leadership. The absolute differences were quite large.

Practice effect. We expected an increase in participation under high motivation as members became more familiar with each other, and a decrease in participation from feelings of apathy and boredom among those low in motivation as the test progressed. The trends were suggestive but failed to attain statistical significance. It was surmised that since each individual test problem was unreliable, comparing performance on the first half of the test with performance on the second half of the test might still reveal a statistically reliable trend. The Type III analysis of variance was recomputed to contrast the results of the first five trials as a whole with the second five. The expected interaction emerged significant at the .01 level of confidence. The trend in means is displayed in Table 3. Highly motivated *Ss* increased while *Ss* of lower motivation decreased in attempted leadership with continued group experience.

Effect of difficulty. It was expected that participation would be greatest where problems were not so easy that little change was required, nor so difficult that everyone withdrew from the challenge. The results failed to support the hypothesis. The group data were examined by the five levels of problem difficulty ignoring practice. An F of 1.72 emerged when the interaction of difficulty and control variance was tested against the error within groups. (An F of 1.90 is significant at the .05 level for 8 and 168 *df.*) Groups with control variance of 0.00 or 1.44 showed less total participation on the easier problems, while groups with control variance of 0.64 did not. Table 4 summarizes the three related analyses of variance.

TABLE 1

PROPORTIONAL NUMBERS OF THE 51 GROUPS AND 255 MEMBERS TESTED UNDER VARYING CONDITIONS

A. Groups

Motivation Level	Control Variance				
	0.00	0.64	1.44	All	Ratio
High	9	9	9	27	9
Medium	4	4	4	12	4
Low	4	4	4	12	4
All	17	17	17	51	

B. Members

Motivation Level	Amount of Control							
	Control Variance 0.00	Control Variance 0.64			Control Variance 1.44		All	Ratio
	1X	1X	2X	3X	1X	4X		
High	45	27	9	9	36	9	135	9
Medium	20	12	4	4	16	4	60	4
Low	20	12	4	4	16	4	60	4
All	85	51	17	17	68	17	255	
Ratio	5	3	1	1	4	1		

TABLE 2

MEAN ATTEMPTED LEADERSHIP IN SECONDS TALKED PER PROBLEM BY THE AVERAGE MEMBER AS A FUNCTION OF MOTIVATION AND CONTROL VARIANCE

Motivation Level	Control Variance			
	0.00	0.64	1.44	All
High	25.2	28.4	34.7	29.5
Medium	28.6	18.9	18.2	21.9
Low	15.1	13.8	22.2	17.0
All	23.6	22.7	28.1	24.8

Analyses of Variance of Individuals Within Groups

The deviation of each member from his own group's mean was calculated for each problem. Since each group mean deviation was set at zero, all group differences, and the *simple* effects associated with motivation, control variance, practice, or difficulty were eliminated

(although their interactions with amount of control were not). The variance in these deviations could be attributable only to amount of control of a member within a group, the interaction of amount of control with motivation, or period, or problem difficulty and the fluctuations between *S*s or within *S*s on successive trials.

TABLE 3

Mean Attempted Leadership in Seconds Talked per Problem by the Average Member as a Function of Group Practice and Motivation

Motivation Level	Amount of Practice: First Phase (Average of first five trials)	Second Phase (Average of second five trials)	All
High	28.2	30.8	29.5
Medium	22.1	21.7	21.9
Low	17.7	16.3	17.0
All	24.3	25.2	24.8

Two separate sets of analyses of variance were run. The first concerned the 17 groups with a control variance of 0.64 and the second concerned the 17 groups where variance in control was 1.44. Table 5 summarizes these two analyses. Amount of control did not vary in the remaining 17 groups where all members drew 1X, hence, these groups could have no remaining variance other than subject error. Nothing further was done with these groups.

As noted in Table 5, the variance estimate due to amount of control was significantly higher than the error between *S*s at the .01 level of confidence in both analyses. The ex-

TABLE 4

Analyses of the Variance of the 51 Groups

Source	*df*	Variance	*F*	Source	*df*	Variance	*F*	Source	*df*	Variance	*F*
Motivation (M)	2	176,192.25	6.67**								
Control Variance (C)	2	31,745.75	1.20								
M × C	4	37,176.38	1.46								
Error between groups	42	25,403.17									
Between groups	50	32,630.29									
Periods (P)	9	2,327.11		Difficulty (D)	4	3,435.50		Phase (Ph)	1	2,823.52	
P × M	18	2,662.53		D × M	8	2,451.05		Ph × M	2	146,084.93	5.48**
P × C	18	3,509.14		D × C	8	4,471.80	1.72	Ph × C	2	42,868.13	
P × C × M	36	1,253.10		D × C × M	16	1,667.50		Ph × C × M	4	38,739.55	
Error within groups	378	3,242.01		Error within groups	168	2,599.40		Error within groups	42	26,739.53	
Within groups	459	3,055.82		Within groups	204	2,610.30		Within groups	51	32,524.47	
Total	509			Total	254			Total	101		

** $p < .01$.

TABLE 5

Analyses of the Variance of the Deviations of Subjects from Their Own Group Means

Source	17 Groups of Control Variance of 0.64: *df*	Variance	*F*	17 Groups of Control Variance of 1.44: *df*	Variance	*F*
Amount of control (A)	2	183,176.85	5.35**	1	497,431.00	11.03**
A × Motivation (M)	4	28,441.78		2	242,491.60	5.38**
Error between *S*s	76	34,534.32		79	45,090.02	
Between *S*s	82	37,862.55		82	55,421.05	
Period (P) × A	18	4,216.36	1.78*	9	5,196.97	1.25
P × A × M	36	2,787.73		18	6,161.12	
Error within *S*s	684	2,345.26		711	4,113.17	
Total	820	5,957.48		820	9,300.80	
Difficulty (D) × A	8	2,695.56	1.21	4	6,408.10	1.70
D × A × M	16	1,538.93		8	7,775.77	2.07*
Error within *S*s	304	2,268.34		316	3,760.60	
Total	410	9,367.05		410	14,196.86	

* $p < .05$.
** $p < .01$.

pected interaction of amount of control and motivation emerged in both situations but only in the more extreme 1X versus 4X condition did it reach significance at the .01 level of confidence. Table 6 shows the mean deviation of *Ss* of varying amounts of control under different motivating conditions. Members of high control attempted more leadership than the others in their own group, particularly if all were highly motivated.

One unexpected result shown in Table 6 was the emergence of a "Vice-Presidential" control effect. Generally, attempted leadership was higher, the higher a member's amount of control, in an absolute sense. The average 1X members in the groups under the two control variance conditions deviated 0.84 and −2.42 seconds respectively from their own group means. The average deviation was 5.10 seconds for the 3X members and 9.68 seconds for the 4X members. However, 2X members, in competition with 1X and 3X members in groups with a control variance of 0.64, consistently attempted the least amount of leadership for all three motivation levels exhibiting an average deviation from their own group mean of −7.62 seconds.

Practice effect. Fig. 1 charts the trends in deviations of *Ss* from their own group means with practice. An F of 1.78 significant at the .05 level was obtained for the interaction of period and status of 1X, 2X or 3X. A corresponding F of 1.25 (not significant) was found for the interaction of practice with 1X and 4X status. Inspection of Fig. 1 suggests that following an initially high rate of attempted leadership by persons of high control there was a drop followed by a rise ending with the 3X members attempting relatively as much leadership as the 4X members. The 4X members did not exhibit such change. It may be the 3X-2X-1X condition was more productive of conflict

TABLE 6

AVERAGE DEVIATION OF SUBJECTS FROM OWN GROUP MEAN IN SECONDS TALKED PER PROBLEM AS A FUNCTION OF AMOUNT OF CONTROL AND MOTIVATION

	Amount of Control				
	Control Variance 0.64			Control Variance 1.44	
Motivation	1X	2X	3X	1X	4X
High	1.61	−10.70	5.88	−4.66	18.66
Medium	−0.72	−6.62	8.78	0.36	−1.42
Low	0.69	−1.70	−0.38	−0.14	0.56
All	0.84	−7.62	5.10	−2.42	9.68

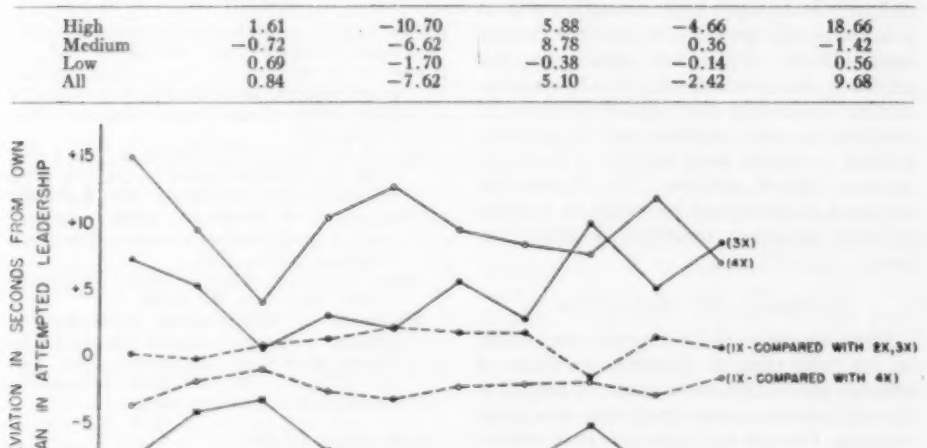

FIG. 1. TRENDS IN ATTEMPTED LEADERSHIP BY SUBJECTS DIFFERING IN AMOUNTS OF CONTROL DUE TO STATUS

and a struggle for status than the 1X-4X situation, but after some experience, the 3X member emerged attempting as much leadership as his 4X counterpart. Although motivation did not interact significantly with the above effects, inspection of the trends under different levels of motivation (not shown) suggested that under high motivation the decline and subsequent rise again of the 3X member was even greater than for all motivation levels combined.

Effects of Difficulty

Maximum attempted leadership by members of high control was expected with problems of average difficulty, problems not so easy that leadership was not required but not so difficult that everyone became apathetic.

As problems became more difficult, there was a tendency for the high control individuals to increase their tendency to attempt leadership relative to that attempted by the members of low control, although, as reported in Table 5, the trends were not significant. However, a significant interaction emerged in the 1X-4X situation. The interaction between difficulty level, amount of control and motivation was significant at the .05 level of confidence. The 4X members attempted more leadership at all difficulty levels under high motivation with an increase in this tendency as problems became more difficult. Under low motivation, this tendency was reversed with the 4X members tending to attempt more leadership than 1X members in easier problems and 1X members tending to display more attempted leadership on more difficult problems. The 3X members increased in attempted leadership as problem difficulty increased, regardless of motivation level.

Summary and Conclusions

Fifty-one groups of five *S*s each were studied on ten brief trials to examine the effects of differing motivation and amount of control in the tendency to attempt leadership in a group situation. On each trial, members rank-ordered judgments before and after group discussion. Groups varied in motivation to perform well according to pretest questionnaires and were assembled so that all members were either high, middle, or low in motivation to be accepted for Advanced ROTC training. Control, the ability to reward or punish other members, also was varied within each motivational group level. Attempted leadership was defined as the total amount of time spent by the *S* in talking in initially leaderless group discussions.

The results point to the significance of motivation and amount of control in the study of behavior in groups. Specifically, they suggest that a member's attempted leadership is higher, the higher his control and the higher the group's motivation. Control yields power only where members are highly motivated. Problem difficulty and group practice must also be considered in understanding the relations between power and leadership. On the whole, less leadership was attempted where problems were easy. With continued practice, highly motivated *S*s increased in attempted leadership while unmotivated *S*s did the opposite. Practice also had complex effects on attempted leadership as a function of control due to status.

References

1. Bass, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.*, 1949, **33**, 527–533.
2. Bass, B. M. Outline of a theory of leadership and group behavior. *Tech. Rep.* No. 1, Contract N7 ONR 35609. Baton Rouge: Louisiana State Univer., 1955.
3. Bass, B. M., Gaier, E. L., Farese, F. J., & Flint, A. W. An objective method for studying group behavior. *Psychol. Rep. Monogr.*, 1957, **3**, 265–280.
4. Lindquist, E. F. Design and analysis of experiments in psychology and education. New York: Houghton Mifflin, 1953.
5. Noble, C. E. The meaning familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.
6. Willerman, B. Organizational involvement as reflected in type of member complaint: An indirect method of measurement. *Tech. Rep.* No. 4, Contract N8 ONR 66216. Minneapolis: University of Minnesota, 1954.

Received May 31, 1957.

EFFECTS OF MOTIVATING INSTRUCTIONS ON REACTION TIME IN CEREBRAL DISEASE[1]

HAROLD L. BLACKBURN

State University of Iowa and VA Hospital, Iowa City

THE influence of manipulable motivational factors on behavior in brain-damaged patients is a topic that has received little systematic investigation. Yet it is a problem with important implications. Among clinical workers, one finds two contrasting points of view (which are not always explicitly stated) regarding the significance of motivational variables in determining the nature and extent of behavioral impairment following cerebral injury or disease. The more traditional point of view conceives of the physical injury as being virtually the sole determinant of the behavioral deficits that are observed to follow the occurrence of the injury, and motivational or other factors are considered as being of little significance as determining variables (10, 15). Moreover, a contrast between this presumably anatomically grounded type of behavioral deficit and apparently similar behavioral deficits encountered in "functional" disorders (e.g., schizophrenia) is often made, the latter being conceived as being significantly if not primarily determined by motivational disturbances.

In contrast to the traditional idea that the physical injury is practically the sole determinant of the associated behavioral deficit in patients with cerebral disease, a more recent point of view conceives of the observed deficits following cerebral damage as being multiply determined and emphasizes the importance of special motivational procedures in rehabilitation work with these patients (4, 13, 18, 19). Standard procedures (orientation to the task with little or no urging or evaluation) are generally considered to be ineffective with such patients. On the other hand, procedures which involve failure experiences and the setting of difficult goals are frequently reported to result in "catastrophic reactions" with consequent severe disruption of performance. "Relaxing" procedures (emphasizing success experiences, reassurance, and minimal frustration) are generally considered to be optimally effective. However, since opinion to date regarding the effects of special motivational procedures has been based on nonquantitative clinical observations of complex and rather ill-defined performances, it seemed desirable to undertake systematic studies of the effects of motivational variables on the performances of brain-damaged patients in such a relatively simple and clearly defined task as the reaction time experiment.

A survey of the literature (1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 14, 16, 17) indicates that considerable doubt exists concerning the type of changes to be expected in the RT performances of either normal or brain-damaged *Ss*, either as a function of simple practice or of special motivating conditions. The present study, which investigated the effects of three kinds of motivating instructions (standard, urging, relaxing) on choice reaction time in nonpsychotic patients with cerebral disease or injury and selected control patients, was designed to answer the following questions: 1. Do brain-damaged *Ss* show changes in speed of choice reaction as a function of either of two types of special motivating instructions? 2. Are these motivational effects comparable in size and direction for both brain-damaged and control *Ss*? Do brain-damaged *Ss* show performance decrements under the standard instruction condition that are not shown by control *Ss*?

METHOD

Subjects and Apparatus

A total of 195 patients from the University Hospital and the Veteran's Administration Hospital, Iowa City, were given two consecutive administrations of a test involving a choice reaction to one of two lights. The first administration was given always under standard

[1] Condensed from a doctoral dissertation submitted to the State University of Iowa and done under the direction of Arthur L. Benton. This investigation was supported by a research grant (B-616) from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service. Sincere thanks are due to Leonard Feldt for valuable statistical guidance and to Russell Meyers, Adolph Sahs, and Maurice Van Allen for their cooperation in obtaining and evaluating the patients used in this study.

instructions. The second administration was given under one of three sets of motivating instructions (standard, urging, relaxing). The patients were assigned on the basis of medical diagnosis to one of two groups: an experimental, or brain-damaged group, and a control group. One third of the patients in each diagnostic group were assigned to each instruction subgroup. In order to control for possible differential effects on the criterion scores of the factors of sex of patient or the facility in which he was hospitalized, the *S*s within each diagnostic group were further divided into subsamples. Assignment of *S*s was so made that the proportions of VA hospital males to University hospital males to University hospital females remained constant for each instructions-diagnosis combination.

The experimental group consisted of 78 nonpsychotic, brain-damaged patients who carried final and unequivocal diagnoses of supratentorial cerebral disease or injury. A variety of diagnoses was represented in the group, vascular disease and neoplasm being the most frequent. The control group consisted of 117 patients from the neurological, neurosurgical, and medical services who showed no evidence or history of cerebral disease or injury. A variety of physical complaints and diagnoses was represented in this group. The mean age of the 78 brain-damaged *S*s was 38 years; their mean educational level was 10 years. The mean age of the 117 control *S*s was 37 years, and their mean educational level was also 10 years. The general criteria for the selection of *S*s and the apparatus employed were identical with those employed in previous studies (2, 3).

Administration

The *S* was seated in front of the light panel of the apparatus. No arm rests were provided. The experimenter sat across the table on the control side of the apparatus. The following instructions were given:

> This is a test to see how fast you can react when you have to make a choice. This time, you place one hand on this key (point) and the other hand on this key (point) and rest them there without pressing down. I'll start a buzzer from over here again; and when the buzzer stops, *one* of the two lights on the panel in front of you will flash on. Keep watching for the light to flash, and when you are sure the right light is on, press down the right key, but *not* the left, as quickly as you can. When you are sure the left light is on, press down the left key, but not the right, again as quickly as you can. Never press down both keys at one time. It's either one or the other. Is that clear? (Repeat relevant portions as necessary). Now we'll try some practice (5 practice trials). That's the end of the practice; now we'll start testing (30 test trials).

Any response completed in .100 second or less, or any depression of the key on the incorrect side, regardless of whether the correct key was depressed before or after, was considered to be an anticipatory error, and an additional trial was given. Only times of correct responses were recorded. Thirty trials involving simple RT always preceded the choice RT procedure. The first 30 choice trials included 15 presentations of each light in a fixed, initially randomly determined order. The trials were given consecutively with no appreciable intervening interval. Following a rest period of approximately two minutes, one of the three following sets of instructions was given to precede a retest administration consisting of three warm-up and thirty experimental trials (15 presentations of each light in the reverse order of the initial administration presentations):

> (*a*) *Standard or task-orienting instructions:* "Now let's try this choice test again. The same instructions apply as before. Let's have a few warm-ups first."
>
> (*b*) *Urging and evaluating instructions:* "Now let's try this choice test again. The same instructions apply as before, only this time, I want to see just how *very* fast you can make the reactions. I've just figured your scores for the last series and your average time was ——. You should be able to do better than that. This time after each reaction, I will tell you in hundredths of a second how long it took you to react. Try to adjust your reactions so as to continually do better. Knowing how you are doing each time should help you. Let's have a few warm-ups first." (After each trial, *E* accurately reports the time and makes evaluative comments. Times equal to the average ± .03 second elicit 'OK'; slower times elicit 'poor' or 'too slow'; faster times elicit 'good' or 'fine.' In each case, *S* is urged to do even better.)
>
> (*c*) *Relaxing and reassuring instructions:* "Now let's try this choice test again. The same instructions apply as before, only this time, I think you would do better if you just *relax*. You've been doing fine so far but just *relaxing* seems to help everybody. Still try to react as quickly as possible but don't concern yourself about how you are doing. Relax as much as possible while you are waiting for the light. Let's have a few warm-ups first." (After practice, assure *S* he is doing fine—even better than before—and should just continue to relax. Repeat assurance after trials 5 and 20 or whenever *S* makes spontaneous inquiry.)

RESULTS

Initial scores. A preliminary analysis of initial test scores (sum of the time measures for the first 30 trials under standard conditions) was performed in order to evaluate any differences in absolute level of performance before instructions treatments were imposed. Significant initial differences ($F = 31.62$, df 1 and 189, $p < .001$) in absolute level were found between brain-damaged and control *S*s, the brain-damaged *S*s being consistently (and, in most instances, markedly) slower than the corresponding control *S*s.

In view of the differences associated with diagnostic group, the effects of sex and hospital residence on initial level were evaluated by performing two separate 3 x 3 factorial analyses of variance, one for the brain-damaged and one for the control *S*s. Within each diagnostic group, a significant difference among the subpopulations ($F = 4.02$, df 2 and 69, for brain-

damaged, $F = 4.72$, df 2 and 108, for control; $p < .05$ in each case) was found. In both groups, the Veterans Hospital males were the fastest, the University Hospital males were next fastest, and the University Hospital females were the slowest.

The same analyses also provided an opportunity to examine the relative size of the differences among the instructional groups which resulted from the random assignment of *Ss* to the several instructions treatment groups. No noteworthy difference in mean initial score was found among the three instructional subgroups of control *Ss*. However, contrary to expectations based on the *random* assignment of the *Ss* to instructions subgroups, an initial difference ($F = 3.63$, df 2 and 69, $p < .05$) was found among these three subgroups of the brain-damaged population. No satisfactory explanation of this finding can be offered except in terms of an unusual chance happening.

To meet the difficulties associated with multiple differences in initial level of performances, an analysis of covariance of the retest scores (sum of the time measures for the second 30 trials for each *S*), with initial scores being used as the control variable, was considered. However, the assumptions for such an analysis could not be met due to a marked heterogeneity of regression within cells.

Gain scores. Since the assumptions for an analysis of covariance could not be met, gain scores (the sum of the 30 time measures *before* instructions treatments minus the sum of the 30 time measures *after* instructions treatments) were utilized as the criterion measure.

Criterion score means and their standard deviations for each of the 18 cells are presented in Table 1. Inspection of this table reveals marked heterogeneity of variance from cell to cell. Application of Bartlett's test for homogeneity of variance statistically confirms the judgment that these variances differ by more than chance ($p < .001$). However, empirical findings, as discussed by Lindquist (12), indicate that the F test of analysis of variance is not very sensitive to any failure to meet the assumptions of normality of distribution or homogeneity of variance. Moreover, means and variances of the 18 cells were not correlated ($r = .01$). Therefore, it was decided that a 3 x 2 x 3 factorial analysis of variance of these criterion measures would permit at least tentative conclusions as to the significance of the various main effects and interactions. In this respect, it should be noted that because of the heterogeneity of variance, the true risk of a Type I error (rejecting a true null hypothesis) is, by the best available estimates, about two to four per cent more than the apparent level of significance.

Three specific null hypotheses were posed with respect to the criterion score means of the treatment populations corresponding to the several diagnostic-instructional categories: (*a*) the effect of instructional differences upon the criterion score means is identical for the two diagnostic groups; (*b*) in the combined diagnostic populations, there is no difference in criterion score means as a function of differences in motivating instructions; (*c*) in the combined instructions populations, there is no difference in criterion score means as a function of differences between the diagnostic populations.

An examination of the subpopulations factor (Veterans Hospital males, University Hospital males, University Hospital females) and its interactions does not bear directly on any of the three hypotheses and is entirely secondary to the purposes of the present investigation. However, restriction of sampling to any one of these subpopulations would have seriously limited both the available pool of brain-damaged *Ss* and the generality of the present

TABLE 1

MEAN CRITERION SCORES (IN .01 SEC.) AND THEIR STANDARD DEVIATIONS BY SUBGROUPS

Population		Instructions Subgroup		
		Standard	Urging	Relaxing
Brain-damaged Subpopulation				
VA males	M	−49	116	25
	SD	69	159	97
Univer. males	M	−22	112	17
	SD	50	53	255
Univer. females	M	−37	172	50
	SD	81	75	66
Control Subpopulation				
VA males	M	24	86	−22
	SD	61	41	76
Univer. males	M	−26	75	10
	SD	76	120	87
Univer. females	M	31	86	69
	SD	63	67	88

findings. Since the statistical tests rendered tenable the assumptions of no subpopulations effect ($F = 1.95$, df 2 and 177, $.20 > p > .10$) on the criterion scores and no double or triple interaction in the populations involving this factor ($F < 1.00$; $p > .50$ in each case), subsequent tests of the experimental factors (instructions and diagnosis) and conclusions with respect to these factors were made possible at more general levels, unrestricted to the individual subpopulations.

The first null hypothesis is, in effect, an assertion of zero interaction between diagnostic category and instructions. There is some question whether the significance of the F test ($F = 3.53$, df 2 and 177, $p \approx .03$) for this interaction should be attributed to an intrinsic interaction in the population or to failure of the data to meet the assumption of homogeneity of variance or both. However, since the empirical studies of heterogeneity of variance indicate that an F of this size is extremely close to a true five per cent level for the interaction hypothesis, it would seem advisable that the effects of instructions and diagnosis be studied separately at each level of the other and that the observed differences be evaluated with respect to error terms appropriate for and computed from each limited segment of the data.

Mean gain scores for each of the six diagnosis-instructions treatment combinations are presented in Table 2. From inspection, it will be noted that marked differences in mean criterion scores among the three instructions treatment groups were found, both within the brain-damaged population and within the control population. Repetition of standard instructions appeared to result in a decrement in performance for brain-damaged *Ss* and a minimal increment for control *Ss*. The introduction of urging instructions appeared to result in marked improvements over initial test scores for both groups. The introduction of relaxing instructions appeared to have no differing effects from repetition of the standard instructions for control *Ss*. The brain-damaged *Ss*, however, showed an apparent improvement in performance under the relaxing condition as compared with their previously noted decrement under the standard condition.

TABLE 2
MEAN GAIN SCORES FOR EACH OF THE SIX DIAGNOSIS-INSTRUCTIONS TREATMENT COMBINATIONS

Diagnoses	Instructions Group		
	Standard	Urging	Relaxing
Brain-damaged	−38	127	28
Control	10	82	9

Separate analyses within each diagnostic group confirmed the significance of the observed main instructional effects ($F = 11.65$, df 2 and 75, $p < .001$, for brain-damaged group, $F = 10.50$, df 2 and 114, $p < .001$ for control group). Analysis of the simple instructional effects showed that the urging instruction differed significantly in its effects from each of the other two instructions (standard and relaxing) within both the brain-damaged and the control populations. The standard instructions and the relaxing instructions did not differ significantly in their effects within either population. With respect to the control *Ss*, the observed difference in criterion scores between the standard and relaxing instructions group was minimal and nonsignificant. On the other hand, the observed difference with respect to brain-damaged *Ss* was relatively large and approached significance at the preestablished level ($.10 > p > .05$).

Under each instructions condition, the brain-damaged *Ss* showed apparently greater changes than the control *Ss*. Under the standard condition, these changes were in opposing directions (*vide supra*), the difference in mean criterion scores for the two diagnostic groups being statistically significant ($F = 7.28$, df 1 and 63, $p < .01$). Under the other two conditions, the amount of improvement among the brain-damaged *Ss* was not demonstrably different from that shown by the control *Ss*, although the observed difference with respect to the urging condition approached significance ($F = 3.18$, df 1 and 63, $.10 > p > .05$ for the urging condition; $F = 0.39$, df 1 and 63, $p > .50$ for the relaxing condition).

In view of the marked differences between the brain-damaged and control *Ss* with respect to initial level of performance, the question arises as to whether the observed differences in gain scores between the two diagnostic groups are a function of these differences in initial level. The question is of further significance in view of the observed differences

in initial level among subsamples in both groups and among the instructional groups within the brain-damaged population.

In order to analyze the relation between initial and gain scores, the *S*s in each of the six treatment combinations were divided into "fast" and "slow" initial performers on the basis of their relationship to the median for that particular cell. The subsamples were pooled for purposes of this analysis. Separate analyses of variance of each of these six differences in criterion scores indicated that none of the observed differences attained statistical significance. Therefore, it may be inferred that if any relationship exists between initial and gain scores, the effect of initial level or subsequent gain scores is of little relative importance when compared with the effects of other factors influencing level of subject performance.

Blocks of trials. The finding of significant effects of instructions for both the brain-damaged and control groups was based on an analysis of gain scores that were derived from total performance over a considerable number of trials. The question arises as to whether these instructional effects are operative throughout the course of performance over the 30 postinstructional trials or only during part of it, e.g., the first few trials. To answer this question, the mean time scores per trial for blocks of 10 trials each for both the preinstructional and instructional conditions and for each diagnosis-instructions treatment combination were computed.

With respect to the preinstructional condition, it was found that the control *S*s showed no evidence of either positive or negative practice effects and that the brain-damaged *S*s showed a slight negative practice effect. With respect to the several postinstructional conditions, the course of performance was very close to a horizontal line in every case, indicating that whatever the instructional effect might have been, it operated fairly constantly throughout the course of the 30 trials.

DISCUSSION

The present study was concerned primarily with the RT responsiveness of nonpsychotic patients with supratentorial disease or injury of the brain to three types of motivating instructions. The performances of those *S*s were compared with those of nonpsychotic medical patients with no evidence or history of cerebral disease or injury, but who were suffering from one type of illness or another. It was believed that, for purposes of an initial study in the area, such *S*s constituted a more appropriate control group than healthy normals in that the possible influence of the factors of illness and hospitalization was controlled.

Initial scores. Significant differences were found in initial level of performance between brain-damaged and control *S*s. As was expected, brain-damaged *S*s were clearly and markedly slower as a group than control *S*s. However, analyses failed to reveal any significant relation between initial and gain scores.

Gain scores. Analysis of gain scores with respect to their relationships with diagnosis and instructions constituted the primary analysis of this study. On the basis of the finding that a significant diagnosis-instructions interaction existed, it was decided that analysis of the simple effects of these two primary experimental factors was most appropriate. With respect to these simple effects, it was found that the two diagnostic groups responded differently to the standard instructions conditions, with brain-damaged *S*s showing a performance decrement postinstructionally and control *S*s a performance increment. This observed difference between diagnostic groups was significant. The direction of these trends in performance with practice is consistent with previous studies and clinical impressions. Moreover, these findings support the suggestion derived from the findings of an earlier study (2) that a larger number of trials than previously employed (60 vs. 30) might reveal a difference in practice trends which would be significant.

Both diagnostic groups showed apparent performance increments as a function of relaxing instructions. The increment shown by the brain-damaged *S*s was apparently greater than that shown by control *S*s, but not significantly so. Among control *S*s, the observed difference in mean criterion scores between the standard and relaxing instructional subgroups was minimal (10 vs. 9) and nonsignificant. Among brain-damaged *S*s, the same difference was sizeable (−38 vs. +28) and approached significance, and the *direction* of

this difference (a decrement under the standard condition and an increment under the relaxing condition) is consistent with clinical contentions that special motivating conditions which assure success and which foster a relaxed attitude in the patient will improve the performance of brain-damaged *S*s and will tend to counteract the usual practice decrements seen in these *S*s under standard conditions. However, since the observed difference was not demonstrably significant, it provides only minimal support for such an idea. In any case, the present findings suggest that such procedures are at least equally as effective as standard procedures.

Urging instructions elicited the greatest improvements in performance, both for brain-damaged and for control *S*s. The increment shown by brain-damaged *S*s was again apparently greater than that shown by control *S*s but not significantly so. Within each diagnostic group, the effects of urging instructions were significantly different from the effects of either of the other conditions. With respect to control *S*s, this finding is consistent with some earlier investigative conclusions (11, 16). With respect to brain-damaged *S*s, the present finding conflicts with earlier reports that attempts to improve the performance of such *S*s through motivational manipulations were unsuccessful (10, 15). Moreover, it is not consistent with the clinical impression that brain-damaged *S*s frequently react catastrophically to failure instructions and the setting of difficult goals. Only 3 (12 per cent) of the 26 brain-damaged *S*s in this instructional category showed performance decrements and, in each case, the decrements were relatively small. Moreover, 4 (10 per cent) of the 39 control *S*s in the same instructional category showed similar decrements, suggesting that brain damage is not the crucial subject variable which determines such decrements. The present data force the conclusions that brain-damaged *S*s benefit from complex instructions involving both failure and the setting of difficult goals and that they benefit from these instructions to at least the same degree and in the same proportions as control *S*s.

Clinical implications and limitations. In general, the instructions conditions were so selected as to reflect or approximate the main motivational procedures employed in the clinic. The standard condition closely parallels the typical test situation in that the *S*s are instructed to do their best at the beginning and no knowledge of results, evaluation, or further encouragement is given during the course of testing. It may be noted that this condition, which is typical of diagnostic test procedures, provided the best diagnostic discriminations and that discrimination was enhanced by including additional trials (i.e., 60 vs. 30).

The relaxing instructions were so selected as to reflect what might be called the current consensus of opinion among clinicians responsible for the retraining of skills in brain-damaged *S*s. *S*s were told that they had succeeded initially. Then the suggestion was made that they would probably do even better. After further attempts, they were reassured that they had met the new goals set for them. Minimal frustration, success experience, and reassurance were the intended major elements of this procedure. Failure to achieve clearly significant effects from such a procedure suggests at least two hypotheses. First, it may be that interfering intensities of concern were not aroused by the type of task employed in this investigation so that procedures designed to reduce concern or anxiety would not be expected to improve performance. Lack of concern in the present task may be related either to insufficient opportunity for the naive *S* to evaluate his own RT performance or to the relatively unimportant nature of the task. This would be contrasted, then, with the usual clinical situation in which the aphasic patient, for example, is fully and painfully aware of his failures and highly conscious of the importance of the skill in which he is deficient. Under such conditions, relaxing procedures might prove to be efficacious.

A second hypothesis which would account for the present findings with respect to relaxing instructions is suggested both by the data and by impressions of the experimenter. The relaxing instructions were associated with the greatest interindividual variability within both diagnostic groups, suggesting that the particular instructions employed may have been ambiguous in character. This suggestion is supported by subjective observations to the effect that many *S*s construed these directions to mean "lie back in your chair and press the

key casually." Other *S*s appeared to understand and to respond to the intended meaning of the instructions.

Whatever the essential differences in the experimental and clinical conditions with respect to relaxing instructions may be, the present data fail to lend any substantial support to current clinical preference for this type of motivational procedure. Particularly is this true when the results with respect to the urging instructions conditions are considered. Admittedly, the urging condition in the present investigation was complex. It included an initial failure experience in that *S*s were told that they should have done better on the first 30 trials. It included factual knowledge of results so that some postinstructional trials led to failure experience and some to success experience. However, for most *S*s in both groups, the predominant experience was success since about 90 per cent of all *S*s improved in performance. Moreover, this condition included the setting of relatively difficult goals ("do better each time") and continual evaluation and urging by the examiner.

In many respects, then, the urging instructions condition involved factors or elements which have been observed or claimed to precipitate catastrophic reactions and severe decrements in performance among brain-damaged *S*s. The present findings that brain-damaged *S*s show their only demonstrable improvement under this condition and that they improve to at least the same degree and in the same proportions as control *S*s are in direct contrast to the clinical observations or impressions and raise some interesting and serious questions concerning the selection of motivational procedures in rehabilitative work with such *S*s. One possible hypothesis, of course, is that the present results are again a function of the nature of the task. It may be that attempts to raise the motivational level of *S*s who are already concerned or anxious about their handicaps (e.g., aphasics or hemiplegics) would lead to performance decrements or disorganized and competing behaviors. Another possible hypothesis is based on the fact that, in the present study, most *S*s achieved relative success in meeting the experimenter-defined goals. It may be that the setting of goals which are beyond the current capacities of the patient would produce interfering responses and consequent decrements in performance. In any case, the present data provide no support for a point of view that warns against too vigorous "pushing" of brain-damaged patients and little support for one that emphasizes the desirability of relaxation and reassurance procedures.

The present study has been concerned with performance on a simple, high speed task, which is minimally affected by simple practice. Discrepancies between the present experimental findings and earlier clinical impressions suggest the need to investigate possible interactions between motivational instructions and the nature of the task. More complex, important, ego-involving, and anxiety-producing tasks might be employed. Moreover, the complex instructions of the present investigation might be broken down into simpler elements which could be studied separately. The influence of knowledge of results, of success experiences, failure experiences, evaluation, reassurance, and urging might then be determined independently of their joint effects and interactions. Some further attempt might also be made to reduce the suspected ambiguity of the relaxing instructions which were employed. Finally, a determination of the effects of these instructional procedures among healthy *S*s might be attempted in order to assess the influence of such factors as illness and hospitalization, the possible effects of which were controlled in the present study.

SUMMARY

The main purpose of the present study was to determine the RT responsiveness of hospitalized, nonpsychotic patients with supratentorial cerebral disease or injury to three types of motivating instructions and to compare their performances with those of a similar group of hospitalized medical patients without cerebral disease. Initially, 30 visual choice RT trials were administered to each *S* under standard conditions. This was followed by differential instructions (standard, relaxing, urging) and by 30 postinstructional trials identical with the initial 30 trials. The criterion score for each *S* was the difference in total time scores between his 30 preinstructional and 30 postinstructional trials.

Marked initial differences in RT between brain-damaged and control *S*s were found.

However, initial differences were not related to criterion scores within any of the six diagnosis-instructions combinations. Practice effects prior to instructions were minimal. Changes over blocks of trials following instructions were also minimal.

Standard instructions resulted in a decrement for brain-damaged *Ss* and an increment for control *Ss*. Both the relaxing and urging instructions resulted in increments for both groups, with brain-damaged *Ss* showing the apparently, but not significantly, greater improvement under both conditions. Within each diagnostic group, the urging instructions were clearly and significantly superior to the other two instructions in effecting improvements in performance. Relaxing and standard instructions were not demonstrably different in their effects within either group. However, there was a suggestion that relaxing instructions might be more effective than standard instructions with brain-damaged *Ss*.

REFERENCES

1. BAXTER, B., & TRAVIS, R. C. The reaction time to vestibular stimuli. *J. exp. Psychol.*, 1938, **22**, 277–282.
2. BENTON, A. L., & BLACKBURN, H. L. Practice effects in reaction time tasks in brain-injured subjects. *J. abnorm. soc. Psychol.*, 1957, **54**, 109–113.
3. BLACKBURN, H. L., & BENTON, A. L. Simple and choice reaction time in cerebral disease. *Conf. Neurol.* 1955, **15**, 327–338.
4. DUNBAR, H. F. *Emotions and bodily changes*. New York: Columbia Univer. Press, 1954.
5. EVANS, J. E. The effect of distraction on reaction time. *Arch. Psychol.*, 1916, **5**, No. 37.
6. FARBER, I. E., & SPENCE, K. W. Effects of anxiety, stress, and task variables on reaction time. *J. Pers.*, 1956, **25**, 1–18.
7. GOLDSTEIN, K. *After effects of brain injuries in war*. New York: Grune & Stratton, 1942.
8. HENMON, V. A. C., & WELLS, F. L. Concerning individual differences in reaction times. *Psychol. Rev.*, 1914, **21**, 153–156.
9. HOVLAND, C. I. The influence of adaptation illumination upon visual reaction time. *J. gen. Psychol.*, 1936, **14**, 436–359.
10. HUNT, J. MCV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. II. New York: Ronald Press, 1944. Pp. 971–1032.
11. JOHANSON, A. M. The influence of incentive and punishment upon reaction time. *Arch. Psychol.*, 1922, **8**, No. 54.
12. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
13. LONGERICH, M. C., & BORDEAUX, J. *Aphasia therapeutics*. New York: Macmillan, 1954.
14. MILES, W. R. The reaction time of the eye. *Psychol., Monogr.*, 1936, **47**, 268–293.
15. NICHOLS, I. C., & HUNT, J. MCV. A case of partial bilateral frontal lobectomy. *Amer. J. Psychiat.*, 1940, **96**, 1063–1083.
16. ROSENBAUM, G. Stimulus generalization as a function of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
17. SMITH, K. U. Bilateral integrative action of the cerebral cortex in man in verbal association and sensorimotor coordination. *J. exp. Psychol.*, 1947, **37**, 367–376.
18. STRAUSS, A. A., & KEPHART, N. C. *Psychopathology and education of the brain-injured child*. Vol. 2. New York: Grune & Stratton, 1955.
19. WEPMAN, J. M. *Recovery from aphasia*. New York: Ronald Press, 1951.

Received June 3, 1957.

AUTONOMIC FEEDBACK: THE PERCEPTION OF AUTONOMIC ACTIVITY[1]

GEORGE MANDLER, JEAN M. MANDLER, AND ELLEN T. UVILLER
Harvard University

DURING the past several years there has been an increasing concern with the measurement and manipulation of the anxiety variable in human subjects. This interest is in contrast to May's summary in 1950: "... the problem of human anxiety has been absent from the strictly experimental specialties of psychology ... (and) ... has largely been omitted from the other branches of academic and theoretical psychology as well." (8, p. 99).

The measurement of anxiety has largely relied on the use of self-rating scales such as the Manifest Anxiety Scale (10) and the Test Anxiety Questionnaire (7). These scales have depended for their construction on some a priori definition of anxiety. In the case of the Manifest Anxiety Scale, this definition leaned heavily on the identification of symptoms consonant with Cameron's specification (2). For the Test Anxiety Questionnaire, the defining operations for item choices were even more vaguely dependent on an intuitive appreciation of the anxiety concept. The point need not be labored that these item choices depend to a large extent on the rather nebulous agreement as to what anxiety "really is." Such an agreement would be easily reached if the anxiety concept were securely anchored in a theoretical framework which provided invariant indices. Such indices are not available for anxiety as an individual difference variable in human subjects. Neither psychoanalytic theory nor the neo-Hullian formulation, to cite the two most frequently used models, provide clear-cut definition for the identification of the variable in human subjects.

This lack of specification for a variable which is subsumed under the emotions is not a novel problem in psychology. In 1890, William James issued a plea for a unified theoretical basis for emotional variables:

> But as far as 'scientific psychology' of the emotions goes, I may have been surfeited by too much reading of the classic works on the subject, but I should as lief read verbal descriptions of the shapes of the rocks on a New Hampshire farm as toil through them again. They give one nowhere a central point of view, or a deductive or generative principle. They distinguish and refine and specify *in infinitum* without ever getting on to another logical level (3, p. 448).

More recently Brown and Farber (1) have suggested a lucid solution for this problem within a Hullian framework. Their formulation, however, is not intended to make the definition of any particular emotional variable more specific. It appears, for the present, to be valuable to specify objective variables which are independent of any general solution for a theory of the emotions. In particular, it is suggested that the concept of anxiety has frequently been used in a global and complex fashion, subsuming many different constituent variables.

The study to be presented here is the first in a series of investigations of one such variable, that of autonomic feedback. Autonomic feedback is defined as the relationship between autonomic response and the subject's reported perception of such response-induced stimulation. While this feedback phenomenon encompasses a variety of different visceral as well as skeletal effects, the major source of stimulation is associated with the activity of the autonomic nervous system. We suggest that this phenomenon represents one variable which has frequently been described as part of the anxiety complex. We will take a position of theoretical neutrality on its specific relationship to anxiety. The study of autonomic feedback may be theoretically useful and empirically productive whether or not it is related to the anxiety concept. However, the three major sources of this concept do stem from general considerations of emotions and anxiety:

1. Autonomic feedback is related to James' position that so-called emotional states can best be described as perceptions of bodily (physiological) activities. It is also related to Lange's (and more recently Wenger's) (4, 11) definition of emotions as patterns of physiological response.

[1] This study was supported by a grant from the Foundations' Fund for Research in Psychiatry. It was reported at the April, 1957 meetings of the Eastern Psychological Association.

2. In the psychoanalytic literature, insofar as it has stayed close to Freud's original formulations, anxiety is primarily a perceptual or ego phenomenon. It is the subject's report of a subjective state, rather than the objective occurrence of any symptoms, which defines anxiety.

3. The anxiety scales mentioned above contain an appreciable number of items related to the subjective report of physiological events.

The first question to be raised concerning the problem of autonomic feedback is a simple one. What is the relationship between the occurrence of autonomic response and the subject's reported perception of such activity?

Physiological events are, in the daily life of the human organism, private events. They stimulate receptors, primarily kinesthetic ones, and are therefore one of the prior conditions for the subject's report of their occurrence. However, the report of the perception of these events is necessarily couched in words which have a different history than do those words which report the perception of public, external events. Skinner (9) has presented several arguments for the nonacceptability of such emotional terms for psychological science; they are primarily based on an analysis of the conditions of reinforcement for such words as "anxious," as compared with the conditions of reinforcement for such words as "red" or "table." Briefly, the contingency between the public event or object and the occasions of reinforcement for the terms applied to them is under the control of the reinforcing community. Such control is not possible for the acquisition of mental or emotional terms. A child will be reinforced for the use of the term "table" whenever in the presence of a table; he may be reinforced for the use of the term "anxious" or "angry" under a variety of internal stimulating conditions which are unknown to the reinforcing community. The result is a wider variability in the use of the latter as compared with the former class of terms.

Under such conditions, it is unwise to expect a simple relationship between the observation of autonomic activity—once made public—and the subject's report of its occurrence. In this initial study we use a twofold approach to the problem of report, and then, given a presumably reliable measure of report, we relate it to autonomic activity.

An examination of the relationship between self-report and autonomic activity is also relevant to the validity of one of the assumptions underlying the anxiety scales mentioned above. It is often tacitly assumed that the report of these subjects is veridical. That is to say, a subject who reports a high degree of subjective anxiety or of symptoms indicative of such anxiety does in fact exhibit these symptoms. However, two different states of affairs may account for a report of considerable internal response: (*a*) The high scoring subject may actually exhibit a greater amount of symptoms or autonomic activity than the low scoring subject, or (*b*) the high scoring subject may not differ from the low scoring subject in reactivity but be more sensitive to and more likely to report small changes in autonomic response.

This study will attempt to deal with this issue by examining the level of autonomic reactivity in two groups of subjects, those who report a high as against a low degree of autonomic feedback.

METHOD

Design. A questionnaire dealing with the frequency and intensity of autonomic self-perceptions was administered to 166 *S*s. The 19 highest scoring *S*s and the 13 lowest scoring *S*s were exposed to an intellectual stress situation. Subsequently, they were interviewed as to their autonomic perceptions during the stress experience and ranked on the basis of this interview. The records of the 14 *S*s who obtained high scores on both the questionnaire and the interview and the 9 *S*s who scored low on both instruments were then examined for level of autonomic reactivity.

Questionnaire construction. The Autonomic Perception Questionnaire (APQ) consists of three sections. The first section requires free response descriptions by the *S*s of their states of feeling and reactions when (*a*) in a state of anxiety and apprehension and (*b*) in a state of pleasure. The *S*s were instructed that when later questions referred to anxiety or pleasure, they should understand by those terms the state which they described in this section. The second section consists of 30 graphic scale items, each of which deals with the perception of bodily activity. These items cover seven areas of bodily reaction: heart rate, perspiration, temperature changes, respiration, gastro-

intestinal disturbance, muscle tension, and blood pressure. An item is composed of such questions as: "When you feel anxious, how often are you aware of any change in your heart action?" with a 14.5 cm scale, the end points of which are marked "Never" and "Always." *S* can mark the scale at any point. For scoring purposes, the scale is divided into 10 equal intervals, and the *S* is assigned a score from zero to nine for each item. Of the 30 self-rating scales, 21 relate to feelings when in a state of anxiety, and nine relate to the state of pleasure.

The third section of the questionnaire consists of 70 MMPI items. Fifty are from the Manifest Anxiety Scale, and an additional 20 were selected from the MMPI as dealing with reports of internal bodily stimulation. These 20 items, together with 14 contained in the MAS, make up a new 34-item Body Perception Scale (BPS).

The questionnaire was administered to 166 Harvard College students (primarily sophomores) as part of a course requirement. Prior to the administration of the questionnaire, *S*s were asked to indicate whether they would be willing to volunteer for subsequent sessions, without being informed of the nature of these sessions.

The stress situation. The *S*s for the stress experiment were selected on the basis of their total score on the 21 anxiety items of the APQ. Of the 25 highest scoring and the 25 lowest scoring *S*s who were requested to participate in another individual session, 19 Highs and 13 Lows kept their appointments. Volunteering did not affect the selection of *S*s and showed no differential tendency in participation.

The stress situation consisted of three tasks with concomitant physiological recording. *S* was seated with his back to the experimenter and the recording apparatus, facing a wall on which the task materials were presented. After a 20-minute adaptation period, *S* was told that he was to be given a series of standardized intelligence tests which should present no difficulty to the average college student. The tasks, a logical reasoning test, a vocabulary test, and a verbal maze, were designed to be extremely difficult, and no *S* was able to complete any of them satisfactorily. They were presented to all *S*s in the same order:

1. The reasoning test consisted of 15 items. *S* was presented with a series of propositions. He was then given 15 statements and was required to indicate the truth or falsity of these deductions from the propositions. The mean number correct on this test was 7.2 (out of 15).

2. The vocabulary test consisted of 10 items, each of which presented a word and a choice of five definitions. The mean number correct was 2.6 (out of 10), which indicates the difficulty of the task.

3. In the Verbal Maze, *S* was presented with a series of choice points consisting of 2, 3, 4, or 5 letters. Each particular combination of letters had one correct choice which permitted *S* to go on to the next choice point. A trial consisted of 12 such choice points, which were repeated for every trial. After four trials, *S* was told that he was doing very badly (which every *S* was in fact) and then was given two more trials.

Physiological measures. A six-channel Grass polygraph was used to obtain measures of autonomic activity.[2] The basic equipment consisted of a Grass pen-recorder, driven by standard Grass D.C. amplifiers. Five channels were used in the present study:

1. *Heart rate.* A cardiotachometer gives either a direct reading of the conventional EKG record, or a tachometer record which supplies a reading of interbeat intervals which can be directly translated into beat-to-beat changes in heart rate.

2. *Psycho-galvanic response.* Preamplification was supplied by a Fels Dermohmmeter,[3] which incorporates an automatic range reset feature. The subject current is 70 microamps. Cup electrodes were used on the *S*'s palms.

3. *Respiration.* A standard pneumograph and a Grass PT-5 pressure transducer were used.

4. *Face temperature.* A Tele-Thermometer[3] with a thermistor probe placed on *S*'s cheek was employed.

5. *Blood volume.* A Waters oximeter, measuring changes in the opacity of the earlobe, was used as a pickup. The major variable contributing to change in opacity is assumed to be blood volume, which in turn is related to pressure changes.

Postexperimental interview. Following the stress situation, the *S* was given a structured interview to elicit his perceptions of autonomic changes during stress. Eight questions tapped areas similar to those used in the Autonomic Perception Questionnaire. The interview was scored by assigning values of two, one, and zero to the *S*s' responses, corresponding to reports of much, little, or no autonomic feedback.

RESULTS AND DISCUSSION

Questionnaire and interview data. Table 1 shows the intercorrelations between the two

[2] Major credit for the instrumentation goes to Albert Grass and Walter Bird of the Grass Instrument Company, Quincy, Massachusetts.

[3] Obtained from the Yellow Springs Instrument Co., Yellow Springs, Ohio.

TABLE 1

INTERCORRELATIONS AND MEANS FOR QUESTIONNAIRE SCALES

	Anxiety APQ		Pleasure APQ		MAS		BPS
Anxiety APQ	—						
Pleasure APQ	.50	.45	—				
MAS	.52	.51	.31	.28	—		
BPS	.52	.52	.20	.37	.67	.73	—
Mean volunteers ($N=74$)	81.6		25.7		17.51		6.46
Mean nonvolunteers ($N=92$)	81.6		22.8		16.34		6.24

Note.—The first correlation given is for volunteer *Ss*, the second for nonvolunteer *Ss*.

parts of the APQ (Anxiety and Pleasure), the MAS (Manifest Anxiety Scale) and the BPS (Body Perception Scale). All correlations, except that between Pleasure APQ and BPS for volunteers, are significant at the .01 level or better. The highest correlation is that of .70 between the MAS and the BPS. When the 14 overlapping items are removed from the BPS, the correlation drops to .60. The hypothesis that the Manifest Anxiety Scale is significantly related to *Ss*' report of internal perceptions is sustained. The Anxiety APQ, the major concern in this investigation, shows consistent correlations around .50 with the three other scales. This finding supports the view that the questionnaire has some communality with other anxiety-related measures. Of special interest are the positive correlations shown by the Pleasure APQ, indicating that *Ss* who report more internal stimulation do so whether the reference is to pleasurable or to anxiety-arousing situations. Whether this is an autonomic or a perceptual phenomenon or a function of response set is, of course, open to question. Finally, there are no significant differences between volunteers and nonvolunteers on either correlations or scale means. This may be related to the fact that *Ss* volunteered before being aware of the nature of the tests to be taken.

As indicated above, the 19 highest scoring and the 14 lowest scoring *Ss* on the Anxiety APQ participated in the stress situation. The range of their postexperimental interview scores extends from 0 to 16, which is the same as the theoretical range. The mean score for the high APQ *Ss* was 7.2; for the low APQ *Ss*, it was 4.1 ($p < .05$). This finding suggests some validity in the APQ, which taps the *Ss*' general perception of autonomic reactions. The specific report of autonomic reactions in a stress situation is related to the more general APQ score.

The interview scores of the 32 *Ss* were divided at the median. Fourteen of the 19 high APQ *Ss* were also high on the interview rating, while nine of the 13 low APQ *Ss* were also low on the interview rating. The 23 *Ss* who showed concordance in their questionnaire and interview scores were selected for the following analysis of autonomic responses.[4]

Physiological findings. For purposes of analysis, the polygraph record was divided into ten reaction periods which represented various aspects of the test process: initial instructions, instruction periods for each test (three periods), and test presentations (six periods). The choice of measures of autonomic activity was dictated by two considerations. The first was previous utility in other studies. Since there are no such measures as the *best* indicator of changes in heart rate, PGR, etc., it was considered to be most efficient to describe the reactivity on any one channel of autonomic activity in terms of more than one measure. The second consideration was dictated by theoretical reasons. Since our primary interest was in autonomic perception, we attempted to select those response aspects which are most likely to be the basis of perceptual discriminations. Such indices as total increase in autonomic response and the variability of responses were considered to fulfill this criterion. The following measures were used:

Heart rate:

1. The mean autonomic lability score[5] for the first and last reaction periods.

2. The variability of mean heart rate over all ten periods. Means were based on the initial 60 secs. of each of the 10 reaction periods.

3. Total raw increase in heart rate from base level to the highest level attained.

PGR:

1. Number of PGR responses during the total record. A response was defined as a change in resistance of 400 ohms or more.

[4] While a study of the discordant *Ss* is certainly of interest, such an investigation was postponed due to the small number of *Ss* in the discordant cells. Further work on this problem is in progress.

[5] The autonomic lability score, developed by Lacey (4), evaluates stress levels corrected for differences in base levels.

2. Mean of the two largest responses occurring during reaction periods two through ten.[6]

3. Total increase in conductance over the entire record.

Respiration:

1. Increase in respiration rate from the base level to the highest level attained.

2. The variability of mean respiration rate based on the initial 60 sec. of each of the ten reaction periods.

3. Change in the standard deviation of inspiration size from reaction period one to reaction period ten.

Temperature:

1. Magnitude of the largest *increase* in temperature regardless of the reaction period in which it occurred. Measured from the base reading of the reaction period.

2. Magnitude of largest *decrease* in temperature, similar to 1.

3. Range over the entire record.

4. Autonomic lability score of the final temperature reading, using the base period as the initial level.

Blood volume:

1. The three largest oximeter responses over the total record.

2. The number of responses over the total record.

3. The variability of response frequency based on ten reaction periods.

Table 2 shows the intercorrelations among these various measures within channels. They indicate that the heart rate, PGR, and temperature measures have the highest intrachannel consistency, while respiration and blood volume measures show much less communality. In general, these correlations confirm the statement that no one measure derived from any autonomic channel is likely to be representative of general reactivity in that channel.

In order to test the major hypothesis concerning differences in autonomic reactivity between high and low perceivers, an over-all rank order was established for each channel. This order was achieved by summing the ranks on each constituent measure and then establishing a new rank order based on these sums. Similarly, a total reactivity rank order was established, based on all five channels. Table 3 presents these ranks for the two groups. The rank difference between the two groups is significant for three channels (heart rate, PGR, and respiration) as well as for the over-all total reactivity score. As expected, the latter shows the best differentiation between the two groups. It should be pointed out that no significant differences were found between the two groups in base levels on any autonomic channel. Thus, *S*s who consistently *report* a high autonomic reactivity also show high autonomic reactivity when compared with *S*s who report little autonomic feedback.

[6] The first reaction period was eliminated since it always included the largest drop in resistance, primarily as a function of shifting *S* from the resting period to the task situation.

TABLE 2

INTERCORRELATIONS AMONG MEASURES WITHIN CHANNELS

Heart rate:	1 vs. 2	.75	Respiration:	1 vs. 2	.49
	1 vs. 3	.84		1 vs. 3	.03
	2 vs. 3	.93		2 vs. 3	.43
PGR:	1 vs. 2	.27	Blood volume:	1 vs. 2	.52
	1 vs. 3	.71		1 vs. 3	.14
	2 vs. 3	.47		2 vs. 3	−.00
Temperature:	1 + 2 vs. 3	.58			
	1 + 2 vs. 4	.32			
	3 vs. 4	.47			

Table 3 also indicates the large variability in ordering from channel to channel. This is consistent with Lacey's previous finding of autonomic response specificity (5). We do not know, however, whether perceptual specificity is related to autonomic specificity, i.e., whether differential report of a particular channel is also related to differential autonomic behavior within that channel. Such a comparison is not possible when extreme *S*s are selected on the basis of *total* APQ scores.

Initially, the question was raised whether *S*s who report a high level of autonomic activity differ in intensity of autonomic response or whether they over-react to autonomic stimulation. Up to this point, the results tend to favor the former interpretation. However, it is possible to evaluate the second hypothesis more directly from the data.

As was indicated earlier, the APQ contains questions relating to all five channels tested in this study. On the basis of their scores on these questions, *S*s can be assigned ranks on the perception of these channels similar to the ranks presented in Table 3. Whenever an *S*'s rank on the *perception* of a particular autonomic channel is higher than his rank on autonomic *response*, he may be said to be overestimating his autonomic reactivity; when these ranks are reversed, he may be thought of as underestimating his reactivity. By this procedure, *S*s can be classified as to their general tendency, i.e., in all five channels, to over- or to underestimate their autonomic activity. Twenty-two of the 23 *S*s could be classified in this manner. Table 4 shows the distribution

TABLE 3
SUBJECTS' RANKS ON FIVE CHANNELS AND TOTAL REACTIVITY

Subjects	Variables					Total Reactivity
	Heart rate	PGR	Temperature	Respiration	Blood Volume	
High perceivers:						
1	9	19	7.5	11	—	13.5
2	5	1	15	—	4	3
3	3	10	2	2	—	1
4	16	14	19	3	3	10.5
5	1	15	16	10	5.5	8
6	21	5	13	5	17	16.5
7	2	9	1	13	2	2
8	—	6	11.5	4	—	5
9	13	4	7.5	12	9	7
10	4	12	21	7	13	12
11	8	—	17	8	11	10.5
12	12	7	5.5	14.5	—	9
13	15	11	4	16	15	16.5
14	6	2	14	6	5.5	4
Low perceivers:						
1	10	20	10	18	1	15
2	17.5	21	5.5	22	14	20.5
3	11	3	3	17	7	6
4	17.5	17	20	1	19	19
5	20	13	11.5	9	18	18
6	19	22	23	14.5	10	22
7	14	8	9	19	8	13.5
8	7	18	18	21	16	20.5
9	22	16	22	20	12	23
Significance of difference between groups (White's T test)	$p < .01$	$p < .01$	$p > .05$	$p < .05$	$p > .05$	$p < .01$

Note.—Missing ranks indicate non-usable polygraph records. In these cases the total ranks are based on prorated estimates.

TABLE 4
AUTONOMIC PERCEPTION AND ESTIMATION TENDENCIES

	Number of *Ss* Showing Discrepancies between Perception and Reactivity	
	Perceptual Overestimation	Perceptual Underestimation
High Perceivers	9	4
Low Perceivers	1	8

of these *Ss* among the high and low APQ groups. The significant chi square ($p < .01$) indicates that high perceivers tend to overestimate their reactivity while low perceivers tend to underestimate. In relation to the initial question, it appears that both processes appear to be operating. *Ss* who report high autonomic feedback not only show a high degree of autonomic reaction but also tend to overestimate that stimulation; similarly, low perceivers show less activity and also underestimate it.

Further studies are in progress to answer some of the questions arising from this investigation:

1. What is the relationship between autonomic feedback and autonomic reactivity for an unselected population, i.e., when dealing with the full range of perceptual and autonomic scores rather than with selected extremes.

2. What is the relationship between task performance and autonomic feedback? In the present study, the tasks were deliberately constructed to be of such great difficulty that no useful individual differences in performance could be observed.

3. What is the relationship between this perceptual-autonomic aspect of the anxiety phenomena and other behavioral indices?

4. Can response specificity be demonstrated in both the perceptual and the autonomic systems?

SUMMARY

Two groups of *Ss* were selected who showed consistency in their report of autonomic activity (autonomic feedback). High perceivers were *Ss* who reported a high level of autonomic feedback in anxiety situations in general and also reported such a high level in a specific stress situation. Low perceivers were *Ss* who reported a low degree of autonomic feedback in both situations. The results showed:

1. Positive correlations between scores on an autonomic perception questionnaire and other paper-and-pencil tests of anxiety reactions.

2. High perceivers showed significantly greater autonomic reactivity than low perceivers.

3. High perceivers tended to overestimate their autonomic responses, while low perceivers tended to underestimate them.

REFERENCES

1. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.
2. CAMERON, N. *The psychology of behavior disorders: A bio-social interpretation*. Boston: Houghton Mifflin, 1947.
3. JAMES, W. *The principles of psychology*. Vol. II. New York: Henry Holt, 1890.

4. LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Annals N. Y. Acad. Sci.*, 1956, **67**, 123–164.
5. LACEY, J. I., BATEMAN, D. E., & VAN LEHN, R. Autonomic response specificity: An experimental study. *Psychosom. Med.*, 1953, **15**, 8–21.
6. LANGE, C. *The emotions*. Denmark, 1885. (Trans. by H. Kurella) Leipzig: Theodor Thomas, 1887.
7. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
8. MAY, R. *The meaning of anxiety*. New York: Ronald Press, 1950.
9. SKINNER, B. F. The operational analysis of psychological terms. *Psychol. Rev.*, 1945, **52**, 270–277.
10. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
11. WENGER, M. A. Emotion as visceral action: An extension of Lange's theory. In M. L. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950.

Received June 6, 1957.

INTRUSION OF ASSOCIATIVE RESPONSES INTO SCHIZOPHRENIC CONCEPTUAL PERFORMANCE[1, 2]

LOREN J. CHAPMAN

University of Chicago

Two traditionally opposed but not necessarily mutually exclusive explanations have been advanced to explain schizophrenic thinking disorder and, particularly, to explain the concept-forming disturbance. The first explanation is that there is a primary loss of ability. The second is that there is what might be called an increased distractibility, i.e., an inappropriate utilization of distracting material which is connected in various ways to the matter at hand. This first explanation is expounded most clearly in the writings of those who view the thinking disorder in terms of the "concrete-abstract" dimension. Prominent in this school is Goldstein (14), who conceives the loss of the "abstract attitude" as the cause not only of the lowered ability to perform on his conceptual tests but also of a wide variety of other thinking aberrations. In his view, inappropriate responses occur by default when the patient does not have available a correct conceptual response.

The second point of view is that the schizophrenic has not lost ability per se, but is prevented from demonstrating his ability by an over-responsiveness to distracting material. Cameron (4, 5) has most vigorously espoused this point of view in his controversy with Hanfmann and Kasanin (15).

Three previous studies (8, 9, 10) have demonstrated that over-responsiveness to various kinds of distracters accounts in part for the lowered performance of schizophrenics. It has been shown that a *differential* over-responsiveness to certain kinds of stimuli rather than others may account for some clinically observable features of the thinking disorder. In each of these studies, there was introduced, as the independent variable, degrees of opportunity to engage in errors of specific kinds hypothesized to be important in schizophrenia. In each case, it was found that on items in which the opportunity to participate in the hypothesized sort of deviant thinking was reduced, the schizophrenics more nearly approximated the performance of a normal control group.

The present study deals with another feature of schizophrenic thinking disorder, inappropriate associative responses. Schizophrenics often appear on clinical observation to be associating to stimuli rather than responding in a more goal-directed manner, as has been noted by well-known psychopathologists (2, 17).

This study offers a task in which errors consisting of response to associative connections may be distinguished from errors of other sorts. On such a task, schizophrenics are expected to make more such associative errors than do normals. In addition, the study tests, for the special case of associative errors, the two traditionally opposed explanations of schizophrenic thinking disorder described above.

The rationale is as follows: If susceptibility to associative distraction is the primary defect, then as the associative strength of distracters increases, the error difference between the groups should increase, at least up to a point. These increased errors should be predominantly associative in nature. On the other hand, if lowered ability to form concepts is the cause of the inferior conceptual performance and of the associative sortings, then the difference between normal and schizophrenic groups on these two aspects of performance should, at least up to a point, increase with a rise in the difficulty of concepts.

METHOD

S was required to sort a card, called here a sorting card, bearing the name of a common object, with one of three guide-card words, choosing the one of the three guide-card words belonging to the same conceptual class as the sorting card word. Each set of three guide-card words consisted of one that shared a concept with

[1] The author is indebted to Julian H. Pathman and George Faibish of the Downey VA Hospital for assistance in obtaining the schizophrenic *S*s, and to Louis Halperin and Meyer Williams of the West Side VA Hospital in Chicago for assistance in obtaining normal *S*s.

[2] The research on which this article is based was done while the author was a postdoctoral research fellow of the National Institute of Mental Health.

the sorting-card word, one which had an associative connection to it, and one which was irrelevant. For example, *S* was required to sort conceptually the word "horse," choosing from among "mule" (conceptual word), "shoe" (associative distracter), and "bottle" (irrelevant word).

Three different bases of associative connection between the sorting card word and the distracter word on the guide card were used:

1. contiguity of objects: e.g., hat-head; raincoat-water
2. contiguity of words: e.g., horse-shoe; gold-fish
3. rhyme: e.g., plate-gate; dancer-cancer

Each sorting-card word was used on various occasions with four different combinations of three guide-card words. The combinations were taken from sets of five guide-card words, such as those shown in Table 1. These four combinations defined the four experimental conditions and were as follows:

1. hard-concept word, high-associative distracter word, and irrelevant word
2. hard-concept, low-associative distracter, and irrelevant
3. easy-concept, high-associative distracter, and irrelevant
4. easy-concept, low-associative distracter, and irrelevant

Since each of these four conditions occurred with the three different bases of associative connection, there were 12 subconditions in all. Thirty items were constructed for each of these subconditions, making a total of 360 items.

Preliminary Development of Test Items

For establishing differences in conceptual difficulty, concepts were offered to judges two pairs at a time (e.g., horse, mule; horse, chicken), and the judges were asked to decide which pair formed the best concept, goodness of concept being defined as the degree to which the two named things are alike in some way. The pair designed to form a relatively poor concept was intended to have an identifiable concept but one less obvious than the pair designed to have a "good" concept.

To establish differences in associative strength, two pairs of words were presented at a time (e.g., horse, shoe; horse, cream), and the judges were asked to decide which pair had the higher associative connection between them, defined as the degree to which one word calls to mind the other. *E* chose the low association words by the same criterion as the irrelevant, i.e., they were intended to have minimal associative connection with the sorting-card words.

The judges were groups of psychology students (28 to 97 per group). On the goodness-of-concept judgments, the median agreement was 98 per cent for the 30 items in each of the three bases of association. In the associative strength judgments, it was necessary to accept differing standards of agreement among the different bases of association. The median percentage agreements for the word pairs were 100 for contiguity of objects, 98.5 for contiguity of words, and 92.5 for rhymes. The high association, low association, and irrelevant categories were roughly equated for frequency of occurrence in the language according to Thorndike and Lorge's (20) word counts.

TABLE 1
SAMPLES OF WORDS USED IN ITEMS

Sorting-Card Words	Guide-Card Words				
	High Association	Low Association	Irrelevant	Easy Concept	Hard Concept
hat	head	street	garage	cap	dress
raincoat	water	food	pencil	umbrella	awning
lip	stick	shoe	chair	tooth	knee
gold	fish	ache	typewriter	platinum	steel
dragon	wagon	handle	rope	sea serpent	trout
face	lace	wagon	weed	toe	liver

Subjects. One normal and one schizophrenic group, each consisting of 32 adult males, were tested. The normal *S*s were medical and surgical patients at the West Side VA Hospital, Chicago, Illinois, for whom there was no indication either in the hospital record or in an interview of psychiatric or neurological disorder. The schizophrenic group, patients in the Downey VA Hospital, had all received a hospital diagnosis of chronic schizophrenia, with their first hospitalization for psychosis occurring at least one year and their last at least 3 months previous to testing. Other prerequisites for selection were absence of any indication of either cerebral organic involvement or mental deficiency, no current drug therapy, no electric shock therapy within the past 3 months, nor more than 35 shock treatments altogether. In addition to the *S*s in the schizophrenic groups, 16 schizophrenic *S*s were also seen but were not included in the final group because they were unwilling or unable to participate.

Mean age was 38.7 years for the schizophrenics and 37.0 for the normals, while mean years of education was 10.7 for the schizophrenics and 10.1 for the normals. Although the groups did not differ significantly with respect to age or education, the schizophrenic group was significantly lower on functioning IQ as determined by the CVS (16), presumably due to the disruptive effects of schizophrenic pathology. On this test, the mean IQ was 108 for the normals, 90 for the schizophrenics.

The *S*s were assumed to be of approximately equal premorbid intelligence because of the equivalent education. An additional approximation of premorbid intelligence was obtained from the item number of the last vocabulary word answered correctly, a measure introduced by Becker (3). This appears to be a better estimate than the total number of vocabulary words answered correctly because one effect of schizophrenia is to produce an unevenness of performance. On this measure, the schizophrenics scored a mean of 8.6, the normals, 9.4. The difference was not significant ($p = .22$).

Although no selection was attempted by diagnostic subtype, the sample was found to include 18 patients diagnosed as paranoid, 9 as hebephrenic, and 5 without classic subtype.

Apparatus. A 13″ x 10″ x 4″ box was used for sorting the cards, the box having three 3¼″ slots along the front edge of the top through which the sorting cards could be deposited. A wire rack for holding the guide cards was placed on top of the box in back of the slots.

The three guide-card words were typed on a 2″ x 11″ card, and were spaced so that when the guide card was placed in the wire rack, each of the three words was in back of one slot. The sorting cards were 3″ x 5″.

Experimental procedure. Before the testing proper, *E* presented six practice items to make certain that *S* understood the nature of the task. These items were identical in form to those in the high association condition of the experiment proper. The only thing that differed was that *E* asked *S* to explain his sorting, and *E* helped him with them. The data from these items were not included in the analyses of results.

After the six practice items, *E* presented 45 items from one of the four main experimental conditions. *E* placed the bundle of 45 guide cards in the rack on top of the box and handed *S* the pack of 45 sorting cards. *E* said, "Now here's a bunch more to sort. For each one, you decide which word up here mentions something that can be called by the same name as the thing mentioned on the card in your hand, and you drop your card in the correct slot, but this time you needn't tell me about it." At the beginning of presentation of the cards for each of the other three conditions *E* again said, "You decide which thing mentioned up here can be called by the same name as the thing mentioned on the card in your hand."

It will be recalled that each sorting card was used with four different guide-card items. Therefore, in order to minimize resultant transfer effects, each *S* responded to only 180 of the 360 sorting card items (15 items in each of the 12 subconditions), receiving the same sorting card in the high-association—hard-concept items as in the comparable low-association—easy-concept items, and similarly receiving the other half of the sorting cards both with the low-association—hard-concept items and with the comparable high-association—easy-concept items. The blocks of cards in each of the four main experimental conditions were presented in a counterbalanced order among *S*s. Within each of the four conditions, items based on the three different bases of association were intermixed in a prearranged random order that was the same for all four main conditions and was constant among *S*s. On the guide cards, words bearing the three kinds of relationship to the sorting word were randomized as to position.

RESULTS

Table 2 shows the total error scores under the four main experimental conditions. In evaluating all the differences, nonparametric techniques were used because the data were extremely skewed.

TABLE 2
MEAN TOTAL ERROR SCORES

Condition	Normals	Schizophrenics
Low Association		
Easy concepts	.09	1.62
Hard concepts	.81	3.38
High Association		
Easy concepts	.66	4.03
Hard concepts	3.09	11.41

As expected, the schizophrenics made significantly more errors than normals ($p < .01$) in all four conditions. This and all other between-group differences were evaluated by a two-tailed Mann-Whitney U test (22).

In addition, manipulation of each of the two independent variables produced significant error differences for each group. With associative strength held constant, both schizophrenics and normals made significantly more errors when concepts became harder, whether associative strength was high or low, and similarly, each group made significantly more errors when the associative strength of the distracters became stronger, both for easy concepts and for hard ($p < .01$ in each case). These and all other within-group differences were evaluated by a two-tailed Wilcoxon test (22).

Of central interest, however, is the amount of divergence of the error curves for the two groups under the various conditions. While this study was primarily concerned with whether the two independent variables are related to *difference* in frequency of errors between schizophrenic and normal *S*s, the effects of the two independent variables cannot meaningfully be compared with one another. However, whether each produces a significant change in errors can be determined.

First, consider the concept difficulty dimension. When associative distracter strength was held low and concepts were varied from easy to hard, the errors of the schizophrenics rose more than those of the normals but not significantly so ($p = .16$). However, when the associative strength was held high, varying the conceptual difficulty in the same way produced a significantly greater rise in errors of schizophrenics than of normals ($p < .01$). Also, a direct comparison of the two divergences revealed that the concept difficulty dimension produced a significantly greater effect in the high association condition than in the low ($p < .01$). Thus, one may conclude that difficulty of concepts increases the inferiority of schizophrenics to normals, but the inferiority is intensified when the test material contains opportunity for incorrect sorting by strong associative distracters.

Next, consider the effects of manipulating strength of associative distracter with difficulty of concepts held constant. For easy concepts, there was a greater rise in errors for schizo-

phrenics than for normals as associative strength increased ($p < .02$). However, for hard concepts, the effect of manipulating associative strength was much greater. This divergence was significant at less than the .01 level, and the difference between the two divergences was similarly significant. Thus, it may be concluded that as the associative strength of the distracters increases, the schizophrenic decrement increases, but the effect of the associative strength dimension is much greater when the correct conceptual solution is more difficult.

Now compare distracter placement errors with irrelevant placement errors for those two conditions under which the associative strength of distracters was held high. Table 3 shows these comparisons, listing the errors separately for the three bases of association.

First, it should be noted that most of the errors consisted of placements with the distracter word, accounting for 89 per cent of the schizophrenics' errors and 92 per cent of the normals' errors in the high association condition.

Errors other than in response to the associative distracter should, by chance, be equally divided among placements with the two incorrect guide-card words. Therefore, to measure errors arising from response to the associative distracter, the number of distracter placements for each *S* was corrected by subtracting from it the number of irrelevant placements. This score of distracter placements minus irrelevant placements should yield a comparatively pure measure of response to the associative connections. Table 4 shows these difference scores for each condition. Because, as noted earlier, the low association and irrelevant words had a similarly low associative connection to the sorting word, distracter-minus-irrelevant scores should approximate zero in the low association condition. Table 4 tends to bear out this prediction.

The analyses which were done for total error score were recomputed, using this distracter-minus-irrelevant score. The results of these analyses were similar to the results reported above for total error scores. On this difference score, each group increased significantly in response to increased difficulty of concepts ($p < .01$ in each case). This means that not only does each group make more total errors when concepts become more difficult, but also, each makes more errors specifically resulting from response to associative distracters. Too, a between-group comparison showed that the schizophrenics significantly exceeded the normals in the extent to which this score rose with an increase in difficulty of concepts ($p < .001$). Thus, as concepts get harder, not only do schizophrenics make more errors than normals; they increase more than normals in identifiable associative errors. In addition, as with total error score, the distracter-minus-irrelevant error score of the schizophrenics diverged more from that of the normals in response to the difficulty dimension when high associative distracters were present than when low ones were present ($p < .01$).

Similar analyses were done to determine the effect of the associative strength variable on distracter-minus-irrelevant error scores. As associative strength of distracters rose, again both groups made higher scores on this difference score just as they did on total errors ($p < .001$ in each case). Again, the schizophrenics exceeded the normals in the extent to which this error score increased as a function of the associative strength of distracters ($p < .001$). Also, the error curve of the schizophrenics again diverged more from that of the normals in response to the associative strength dimension when concepts were hard than when they were easy ($p < .01$).

TABLE 3
MEAN DISTRACTER AND IRRELEVANT ERROR SCORES OF NORMAL AND SCHIZOPHRENIC *Ss*, SORTING ON THREE BASES OF HIGH ASSOCIATIVE DISTRACTION

Association Basis	Normals		Schizophrenics	
	Dist.	Irr.	Dist.	Irr.
Contiguity of objects	2.03	.09	6.75	.31
Contiguity of words	1.03	.09	5.03	.62
Rhyme	.38	.12	1.91	.81

TABLE 4
MEAN DISTRACTER-MINUS-IRRELEVANT SCORES

Condition	Normals	Schizophrenics
Low Association		
Easy concepts	−.06	.12
Hard concepts	−.12	.21
High Association		
Easy concepts	.59	3.53
Hard concepts	2.53	8.41

In conclusion, both independent variables had a strong effect on associative responses, and the two strongly interacted. It is clear, then, that both difficulty in forming concepts and susceptibility to associative distraction are in part responsible for the fact that schizophrenics show more associative intrusions than do normals. Some substantiation is found here for both hypotheses.

DISCUSSION

A supplementary inference from these data is that some aspects of the schizophrenic thinking deviancy are exacerbations of normal response tendencies. The associative distracters, responded to by the schizophrenics, were invented by normals and were judged as having associative connection by other normals. The errors of the normal *Ss*, though fewer in number, were, like those of the schizophrenics, predominantly associative rather than irrelevant placements. Also, the normals, like the schizophrenics, increased their associative choices in response to each of the two independent variables. Thus, the schizophrenics differed from the normals on this task primarily on the frequency of these responses.

Similarly, in each of three previous studies (8, 9, 10), the specific kind of error which accounted primarily for the difference between schizophrenics and normals tended consistently to be committed more often by normals than other types of errors which had an equal opportunity to occur by chance.

It has been suggested by various writers that the thinking errors of schizophrenics resemble those made by normals under this or that special condition. Among the special conditions mentioned have been sleep (13, 17), relaxed attention as in free association (17, p. 17), drugs (18), and fatigue (5, 12).

Cameron and Magaret (6) also reported a finding that normals in completing incomplete sentences in an ordinary classroom situation showed kinds of errors that seemed to resemble schizophrenic thinking disorder. However, when they discuss in detail elsewhere (7) the broader issue of the many similarities of normal errors to those of schizophrenics, they emphasize the importance of special conditions in producing such effects in normals. Among their special conditions are the interruption of ongoing activity, environmental changes, preoccupation, and emotional excitement.

The present findings suggest that the error patterns of normals under ordinary functioning are found exacerbated in what is called schizophrenic thinking disorder. There is supporting evidence for this hypothesis in the data of other investigators. Fey (11) found that on the Wisconsin Card Sorting Test, the perseverative errors most often made by normals were also the kind primarily responsible for the fact that the schizophrenics erred more than the normals. This is especially impressive, considering that the test is constructed so that if responses occurred in a random fashion the perseverative errors would be a little less than half as frequent as the nonperseverative errors.

Further evidence for this kind of relationship is found in the apparent similarity between other solution preference patterns or "biases" of normals, and well-known features of schizophrenic thinking pathology. An example is seen in errors in syllogistic reasoning. Although Von Domarus (21) and Arieti (2) state that a distinctive feature of schizophrenic thinking is a tendency to conclude that two subjects are identical on the basis of their sharing common qualities, Sells (19) and Wilkins (23) have reported that the same error tendency is strong among normals.

Another well-known schizophrenic error pattern is the interrelating or combining of ideas in an inappropriate way. This is especially prominent in delusional patients. However, Abel (1) reported that when her normal *Ss* attempted to reproduce a difficult passage of scientific prose, about half of them produced sentences in which two or more ideas were inappropriately combined into a whole so that the original meaning was lost.

These latter two similarities between normal and schizophrenic thinking are only suggestive because they have not been measured on the two populations by the same tasks. An important direction for new research would seem to be the designing of tasks measuring different types of error preferences to determine whether there are valid correspondences between normal and schizophrenic errors. Such relationships might be sought for facets of the schizophrenic thinking disorder not readily referable to already isolated normal biases. Also, one might determine the extent in schizophrenia of

previously isolated normal biases which are not obviously present in clinical schizophrenic psychopathology.

Clinically, schizophrenics appear to differ greatly among themselves in their thinking disorder. It would be valuable to determine whether such differences may be described as exacerbations of different normal biases, or whether some characteristics of their thinking disorders are not found in normal *S*s.

SUMMARY

The present study tested the hypothesis that schizophrenics' substitution of associative responses for correct responses is due to heightened susceptibility to associative connections, and also tested the alternative hypothesis that it is due to a primary loss of ability to perform correctly. A conceptual card-sorting task was used in which for each sorting choice there were three alternatives, one of which was the correct conceptual response, one an associative distracter, and one neither. The difficulty of the correct conceptual task was varied at two levels, testing the hypothesis that difficulty in forming concepts is responsible for the schizophrenics' greater number of errors, in particular their associative errors. Similarly, there were two levels of the associative strength of the distracter, testing the hypothesis that susceptibility to associative distraction, independent of difficulty of the correct task, is a source of these errors.

Both hypotheses were supported in that both stimulus variables were found related to frequency of associative errors for both schizophrenics and normals, with the divergence of the error scores of the two groups resulting from a greater susceptibility of the schizophrenics to each of the two independent variables. The implications of these results were discussed in terms of the suggestion that some aspects of schizophrenic thinking pathology consist of exacerbations of normal response tendencies.

REFERENCES

1. ABEL, THEODORA M. Unsynthetic modes of thinking among adults: A discussion of Piaget's concepts. *Amer. J. Psychol.*, 1932, **44**, 123–132.
2. ARIETI, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.
3. BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.
4. CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
5. CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. II. New York: Ronald, 1944.
6. CAMERON, N., & MAGARET, ANN. Experimental studies of thinking. I. Scattered speech in the responses of normal subjects to incomplete sentences. *J. exp. Psychol.*, 1949, **39**, 617–627.
7. CAMERON, N., & MAGARET, ANN. *Behavior pathology.* Cambridge, Mass.: Riverside, 1951.
8. CHAPMAN, L. J. Distractibility in the conceptual performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 286–291.
9. CHAPMAN, L. J. The role of type of distracter in the "concrete" conceptual performance of schizophrenics. *J. Pers.*, 1956, **25**, 130–141.
10. CHAPMAN, L. J., & TAYLOR, JANET A. The breadth of deviate concepts used by schizophrenics. *J. abnorm. soc. Psychol.*, 1957, **54**, 118–123.
11. FEY, ELIZABETH T. The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. *J. consult. Psychol.*, 1951, **15**, 311–319.
12. FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
13. FREUD, S. The unconscious. *Collected papers.* Vol. IV. London: Hogarth Press, 1953.
14. GOLDSTEIN, K., & SHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
15. HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, No. 67.
16. HUNT, W. A., FRENCH, ELIZABETH G., KLEBANOFF, S. G., MENSH, I. N., & WILLIAMS, M. Clinical possibilities of an abbreviated individual intelligence scale. *J. consult. Psychol.*, 1948, **12**, 171–173.
17. JUNG, C. G. The psychology of dementia praecox. *Nerv. ment. Dis. Monogr. Ser.*, 1936, No. 3.
18. RINKEL, M., HYDE, R. W., & SOLOMON, H. C. Experimental psychiatry. III. A chemical concept of psychoses. *Dis. nerv. System*, 1954, **15**, 259–264.
19. SELLS, S. B. The atmosphere effect: An experimental study of reasoning. *Arch. Psychol.*, 1936, No. 200.
20. THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words.* New York: Columbia Univer., 1944.
21. VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. Calif. Press, 1944.
22. WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Henry Holt, 1953.
23. WILKINS, M. C. The effect of changed material on ability to do formal syllogistic reasoning. *Arch. Psychol.*, 1928, No. 102.

Received June 7, 1957.

THE RELATIONSHIP BETWEEN JUDGMENTS OF A COMMUNICATOR'S POSITION AND EXTENT OF OPINION CHANGE[1]

WALTER WEISS

Boston University

In some situations, persons to be exposed to a communication may entertain prior expectations concerning the communicators' position that can influence their interpretation of the communication, or its effectiveness. Such anticipatory categorization may result when the audience has had past contact with the communicator or with a similiar class of communicators. But in many circumstances, it is on the basis of the content of the communication itself that the audience comes to view the source as holding a particular opinion position. This judgment, whether vaguely or precisely formed, is presumed to be related to the extent of opinion change exhibited by members of the audience. In the extreme case, misinterpretation of the sense of the communicator's arguments may lead to no change or to a negative effect. But when the general direction of the communicated opinion position is obvious and is opposed to that held by the audience, is there a relationship between the extremeness of the opinions attributed to the communicator and extent of opinion change?

There is some evidence of a positive relationship between extent of opinion change and extent of discrepancy between the communicator's position and that of the initial position of the recipient (1, 3, 9). But as Hovland and Pritzker (4) point out, this naturalistic association may be due in part to a "ceiling effect" (i.e., those whose initial opinions are least discrepant from the communicator's position can change least) or to differential regression toward the mean on the retest. In their own study, amount of change advocated was made independent of the *S*s' initial positions and the phenomenon was confirmed. However, an elaborated, persuasive communication of the kind usual in a "natural" setting was not employed. Accordingly, judgment of the communicator's position was not a variable in their research.

[1] The research reported in this article was supported initially by funds from the Office of Naval Research, under Contract NONR-492(04), and completed under grant M-1351 from the National Institute of Mental Health.

Procedure

An after-only arrangement was employed, involving three groups: two exposed to a communication on juvenile delinquency and a control group that was not. One experimental group read an appeal by an *anonymous* communicator for the adoption of a decidedly, though not extremely, punitive policy toward delinquents. Arguments supporting this position were essentially paraphrases with elaborations and suitable changes of four statements from Form B of the Wang-Thurstone scale of attitude toward the treatment of the criminal. These statements ranged closely around a median value of 8 (11 being the extreme punitive end of the scale). The second experimental group read the same basic communication but with the addition of a few introductory statements to the effect that some people hold to the opinion extremes of complete leniency or excessive punitiveness toward delinquents. An opinion statement was given illustrating each of these end anchors for the opinion continuum.[2] Then the communicator presented his own position, which was identical in content to the material read by the other experimental group.

After reading the two-page article, the *S*s responded to a modification of Form A of the Thurstone scale *as though they were the communicator*; that is, they signified agreement or disagreement with each of the 20 statements as they thought the communicator would, now that they were aware of his position. Following this, they answered a modification of Form B of the scale in terms of their *own* opinions on the problem of combatting delinquency. A series of general judgmental items completed the questionnaire.

The no-communication control group rated

[2] The expectation was that the extreme opinion anchors would induce a contrast effect (6, 9) or extension of the scale of judgment (7, 8), with a consequent decrease in the perceived extremeness of the communicator's position. In answer to a direct question, over 80 per cent of the students in each group did consider the opinion statements as representative of the extremes of the opinion continuum.

each of the statements of Form A on a 1–11 basis in terms of its expressed degree of approval or disapproval concerning the use of punishment. Then members of this group gave their own opinions on delinquency by responding to Form B. They also answered a few judgmental items.

Booklets covering the three treatments were arranged in a random sequence, following a table of random numbers. They were handed out individually to the *S*s who were seen during regular class periods at Boston University and American International College.[3] The purpose of the study was stated to be an examination of people's judgments of the opinions of others on a controversial topic.

RESULTS

The Wang-Thurstone scales were scored by assigning a value of one to a response favoring punitiveness toward delinquents and a zero to a response opposing punitiveness. Neutral items on each scale could not be used under this system. Accordingly, data on the judged position of the communicator are based on responses to 18 items of Form A, and data on the *S*s opinions are based on 17 items of Form B.[4]

When judgments of the communicator's opinion position are examined, it is found that the group that is provided the end anchors locates him at a significantly *less* punitive position than does the other communication group (10.67 vs. 13.30; $p < .001$[5]). This result is supported by congruent and significant differences between the two groups in their responses to two specific questions concerning the author's views on the effectiveness of punishment and guidance in treating delinquents. Thus, judgments of the communicator's position were affected in the desired manner despite identity of the basic facts and arguments supporting and elucidating his viewpoint.[6]

TABLE 1
MEAN AFTER-OPINIONS ON DELINQUENCY

	Communication Groups		
	No standards	End standards	Control group
Mean	6.18	5.22	4.29
SD	2.77	2.53	1.70
N	60	59	68

Analysis by *t* tests of the opinion data presented in Table 1 indicates that the communication group *not* provided the end standards exhibits a significantly *more* punitive after-mean than does the other experimental group ($p = .05$). Thus, it is the group exhibiting the more extreme judgment of the communicator's opinion position that shows the greater effect of the communication. The after-mean of each experimental group is significantly more punitive than that of the control group ($p < .03$ and $< .001$).

The difference between the judged position of the communicator and the opinion mean of the control group can be conceived of as an indication of the amount of change advocated. Thus, for the group provided the end standards, the amount is 6.38 and for the other group 9.01. Using the control mean as the baseline, the amount of change exhibited is 0.93 and 1.39 respectively. If average amount of change is divided by amount advocated, an index of relative change can be obtained. For the group without the end standards, this is 21.0 per cent of the amount advocated and for the other group only 14.6 per cent. (The same result is obtained when the Wang-Thurstone scale values are used.) These data, though crude, do not agree with the finding of Hovland and Pritzker (4) of a negative relationship between relative change and amount of change advocated and obtained.

Data concerning the relationship between judgments of the communicator's position and the *S*'s after-opinions on delinquency are presented in Table 2. By analysis of variance, the main effect of judgment of author's position and the interaction effect were not sig-

[3] The author wishes to thank Malcolm Klein, of Boston University, and Richard Sprinthall, of American International College, for distributing the booklets to their respective classes.

[4] None of the results is changed if the items are scored according to the original scale values given by Wang and Thurstone for attitude toward the treatment of the *criminal*.

[5] All p values are two-tailed.

[6] When the *S*'s judgments of the author's opinion position are scored in terms of the original Wang-Thurstone scale values, the median for the group given the end standards is 6.4 and that for the other group is 7.3. Since the four statements embodying the communicator's position had a median value of 8.0, the group that was not provided the end standards seems to have judged its position more "accurately" than did the other group.

TABLE 2
RELATIONSHIP BETWEEN *Ss*' OPINIONS ON DELINQUENCY AND THEIR JUDGMENTS OF AUTHOR'S POSITION

Author's position	*Ss*' Opinions	
	No standards	End standards
Above mdn.	6.51 (41)	4.86 (21)
Below mdn.	5.47 (19)	5.42 (38)

TABLE 3
RELATIONSHIP BETWEEN AFTER-OPINIONS ON DELINQUENCY AND JUDGMENTS OF COMMUNICATOR'S TRUSTWORTHINESS

	No standards	End standards
Relatively trustworthy	6.86 (22)	5.45 (38)
Other	5.79 (38)	4.75 (20)

nificant. Also, the judgment subgroups within each experimental group were not significantly different from each other. It is only among those who locate the communicator above (i.e., more punitive than) the median judgment of the total group that the communications have significantly different effects (*t* test p value .05).

In response to a multiple choice question, a significantly higher proportion of people given the end standards rated the communicator as relatively trustworthy than did those in the other experimental group (65.5% vs. 36.7%; $p < .01$). Analysis of the data in Table 3 by analysis of variance indicates that the main effect of judgment of trustworthiness and the interaction effect were not significant. Differences between subgroups within a communication group were also not significant.

Judgments of trustworthiness were found to be negatively related to judgments of the communicator's opinion position. That is, those who viewed the communicator as relatively trustworthy judged him to be significantly less punitive than did those who did not consider him to be trustworthy. When the judgments of communicator trustworthiness and opinion position are combined, it is only between those subgroups rating him as relatively trustworthy *and* judging his position to be above the median for all experimental *Ss* that the difference between the two treatments is statistically significant. When either judgment of trustworthiness or of opinion position is held constant, the other judgment is still not associated significantly with the after-opinions of the *Ss*.

Two specific, multiple choice questions sought the *S*'s own judgments of the effectiveness of guidance and punishment as methods for dealing with delinquents (i.e., with the instrumental value of such procedures). Within each group and for the aggregate of all *Ss*, responses to these items were significantly and positively related to the punitiveness of after-opinions. However, the two communication groups did not differ significantly in the proportions judging the instrumental value of guidance or punishment to be high or low. When *Ss* in the communication groups are separated in terms of their responses to these two questions and comparable subgroups are compared, the differences in after-opinions on delinquency are in the same direction as the over-all difference between the two experimental groups.

DISCUSSION

The finding of greater opinion change for the group attributing a more extreme position to the communicator agrees with previous research relating extent of change to discrepancy between initial opinion and advocated opinion (1, 3, 4, 9). However, this judgment dimension does not seem to be related in any simple or obvious manner to the difference in effectiveness of the two communications. This is evidenced by the lack of an over-all or within-groups association between extremeness of judgment of the communicator's opinion and punitiveness of own position, and by the finding of a significant difference in communication effectiveness only between the subgroups judging the communicator's position to be quite extreme. The methodological defect of an artificial restriction or "ceiling effect" on extent of change was avoided in this study; for the position attributed to the communicator by the less influenced group was considerably more extreme than its own mean opinion and than that of the other experimental group.

Internal analyses have not yielded a satis-

factory explanation for the main opinion results. The availability of information concerning the before-opinions of the *S*s and concerning their interpretations of the significance and force of the arguments might have clarified the meaning of the data. As sheer speculation, however, it is conceivable that the method employed for affecting the perceived extremeness of the communicator's position reduced the impact of the communication; that is, mention of the extreme opinion standards may have reminded the students of their own nonpunitive position or fostered a crystallization of opinion, thereby increasing resistance to change.

Hovland and Pritzker (4) have suggested that with an anonymous source, a large discrepancy between the position advocated and that held initially by the audience may diminish regard for the source. This supposition is supported by the current data. But the further expectation that such lowered regard would be reflected in lessened communication effectiveness is not confirmed. It is possible that the condition of anonymity of the communicator changed the personal meaning and functional significance of a judgment of trustworthiness. However, even with an identified and known communicator, there is some research evidence that apparent regard for the source need not be a critical factor underlying certain judgments of a communication or its effectiveness (2, 10).

In general, more information is needed on the psychological *processes* underlying opinion change, so that the influence of particular communication variables and the functional significance of relevant judgments can be fitted within and predicted from the framework of a systematic theory.

SUMMARY

The purpose of the research was to examine the relationship between judgments of a communicator's position and extent of opinion change. Three groups were formed: two exposed to a communication on juvenile delinquency and a control group that was not. One communication group read an appeal by an anonymous author for the adoption of a decidedly, though not extremely, punitive policy toward delinquents. The other experimental group read the same basic communication with the addition of a few introductory statements reminding the audience that some people hold to the opinion extremes of complete leniency or excessive punitiveness toward delinquents. Both groups then responded to a modification of Form A of the Wang-Thurstone scale of attitude toward the treatment of the criminal as they thought the communicator would, and then answered Form B in terms of their own opinions on the problem of delinquency. A series of general judgmental items completed the questionnaire. The control group rated each statement of Form A on a 1–11 scale in terms of its expressed degree of approval or disapproval concerning the use of punishment and then gave their own opinions on delinquency by answering Form B.

The group given the extreme opinion standards judged the communicator's position to be less punitive than did the other communication group. But it was the latter group that was more influenced by the communication. However, when the data for each experimental group and for the combined aggregate of experimental *S*s were examined, judgment of communicator's position was not found to be related to own opinion on the treatment of delinquents. The mean opinion position of each experimental group was more punitive than that of the control group. Although the judgment of trustworthiness of source was negatively related to extremeness of judgment of the communicator's position, it was not significantly associated with the *S*s' opinions.

REFERENCES

1. EWING, T. N. A study of certain factors involved in changes of opinion. *J. soc. Psychol.*, 1942, **16**, 63–88.
2. FINE, B. J. Conclusion-drawing, communicator credibility, and anxiety as factors in opinion change. *J. abnorm. soc. Psychol.*, 1957, **54**, 369–374.
3. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.
4. HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.
5. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
6. MCGARVEY, HULDA R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, N. Y., 1943, **39**, No. 281.
7. ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol.*, N. Y., 1941, **37**, No. 261.

8. SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. Unpublished manuscript. Yale Univer., 1956.
9. SIMS, V. M. Factors influencing attitudes toward TVA. *J. abnorm. soc. Psychol.*, 1938, **33**, 34–56.
10. WEISS, W. Opinion congruence with a negative source on one issue as a factor influencing agreement on another issue. *J. abnorm. soc. Psychol.*, 1957, **54**, 180–186.

Received October 10, 1957. Prior publication.

CASE REPORT

WAS LARS SEXUALLY ASSAULTED? A STUDY IN THE RELIABILITY OF WITNESSES AND OF EXPERTS[1]

ARNE TRANKELL

University of Stockholm

LARS, a five-year-old boy, lived with his parents on the outskirts of a fair-sized town. The family lived in an old house, without the modern conveniences of the newly built apartments nearby. One day, Lars was left alone in the street for half an hour while his mother and grandmother visited the local stores. When they came home, they met Lars just outside the door. He had a bundle of newspapers under his arm and was eating a bit of liquorice that had stained his lips black. When his mother asked him where he had got the papers and the candy, Lars replied that he had got the papers from a "window-cleaner man." As for the candy, he said he had bought it with a big black penny, which he had also had from the window cleaner. On hearing this disturbing tale, the two women drew the boy into the entrance, and there he told them that he had also had to suck the man's penis. They all quickly went up to the apartment, where the grandmother discovered that the boy's fly buttons were undone. This fact, too, the boy explained as a consequence of his encounter with the window cleaner.

WHAT HAPPENED? TWO VERSIONS

According to what the parents told the police, the mother and father agreed over the telephone to go into the matter no further until the police had had a chance to hear the boy. The police report states that the mother visited the police station the same evening and reported the incident. It does not emerge from the report whether the child went with her to the police. A number of the boy's words are, however, reproduced as direct speech in the police report. The most important is the following: "I had to suck his doodle, and he pushed it down so that he nearly choked me. It was so disgusting." The boy said, moreover, that the man lived in the same house as Sune Olsson.

If one follows the police report, one finds that Lars was escorted to the scene of the crime as given by the mother. He pointed out a house some blocks away from his own home. The police quickly established that there lived in this house a window cleaner of foreign nationality, possibly of Russian origin. It also turned out that the window cleaner intended to travel the following day to an unknown destination, so that it was arranged to pick

[1] *Editor's Note.*—Expert evaluation of children's testimony has a long tradition in European forensic psychology. In Sweden, where the case reported by Professor Trankell originates, the courts have come to rely on special experts when the evidence for a charge depends mainly on a child's testimony, as it typically does in cases of assault against minors, usually committed under circumstances that prevent anyone but the accused and the purportedly assaulted child from knowing just what has occurred. If the expert finds the child to be a reliable witness, a conviction may be obtained on the basis of the child's evidence.

The case of Lars, which has the fascination of a good detective story, may serve to acquaint American psychologists with this sort of court service with which they have little if any direct experience. As a case in abnormal and in social psychology, however, its primary interest lies in another direction. Not only does Professor Trankell provide us with new illustrations of familiar phenomena concerning the unreliability of witnesses; what is novel is the way in which his account brings to our attention the more serious problem of reliability that arises when experts disagree.

It is Professor Trankell's suggestion that the experts' divergent findings in the present case follow from divergent methods of approach. One of the experts draws his conclusions from a personality appraisal of the reliability of the child witness. Such an approach is invited by the legal formula that is used in Sweden, according to Professor Trankell, when specialists are appointed: "The Municipal Court has appointed you as a specialist in the case of ____ to give your opinion on the reliability of the ___-year-old child ____." The other expert, Professor Trankell himself, arrives at very different findings from an approach that he characterizes as focussed on the social relationships and interactions of the alleged victim and criminal, attempting to assess not so much the reliability of the witness as the reliability of the crucial statements upon which the case depends. To what extent is the conclusion of each expert the result of his method? The reader may well read the case with this question in mind, remembering that the report is presented by one of the parties to the disagreement.

him up quickly. The window cleaner, a man of 24 here called S., was arrested by the police at about five P.M. at the office of the firm that employed him. Earlier the police had arranged with commendable speed to have the child's reliability tested. The examination took place between two and four P.M. on the same day that S. was arrested, that is, some 24 hours after the event. The test consisted of a conversation between the child psychiatrist and Lars' mother to establish the background, together with a tape-recorded interview with Lars. According to a statement in the report, there was also a summary test of the child's intelligence. The whole investigation took place at the police station, where the mother and the boy had been called.

Besides the child psychiatrist's interview with Lars, the police report included several interviews with S., S.'s fiancée, her mother, the landlord's agent in the building, the assistant in the sweetshop where Lars bought his liquorice, and S.'s employer. A summary of the police enquiry gives the following picture:

S. stated that he had cleaned the windows of a chemical business in the street where Lars had been left alone by his mother and grandmother. When he had finished and was going to cycle on to his next place, he had been accosted by a vagrant, who had begged him for some money. He remained standing in the street for a moment, with the ladder under his arm and the bicycle beside him, while he argued with the beggar. During this discussion a little boy came up to him and asked: "May I carry your ladder, Mister?" S. claimed to have answered that the ladder was too heavy. S. had not observed the boy's appearance closely. When S. had got rid of the vagrant he pushed the bicycle on to the sweetshop, where he had a cleaning contract. He left the bicycle outside the shop on the still unfinished pavement and leaned the ladder against the wall at the side of the shop. Then he went on another hundred yards or so to his own house. He intended to have a break for coffee in his home before cleaning the shop's windows.

Lars' information when he was taken to the place agreed well with S.'s. The boy independently pointed out the place where S. had laid down the bicycle and the place where he had stood the ladder. He also indicated an entrance to S.'s house, through which he claimed to have passed on his way up to "the man's house." The staircase on which S.'s flat was situated was on the inner side of the courtyard inside the long block and could not be reached from the street except through the cellar, unless one wanted to go round the whole block, which meant a detour of several hundred yards. Interrogation of the landlord's agent established that the cellar doors were always kept locked. A little child could therefore not get through the cellar by himself.

S. definitely asserted that he had gone through the cellar alone and that he had unlocked the door and locked it again behind him when he passed through. He denied having heard anyone come after him through the cellar. When he came up into his flat, which was empty during the day, he made some coffee and a few sandwiches. While he was drinking his coffee, there was a ring at the door. When he opened the door he found a little boy on the stairs, who asked if he might have some old newspapers. S. left the door ajar and went into the kitchen, where he took out a bundle of papers in the cupboard above the refrigerator where such things were kept. He came back into the hall and gave the papers to the boy. After that the boy went off. S. formally denied giving the boy any money. When asked if it was the same boy who had spoken to him earlier in the day, S. could not say for certain. He thought that it was quite possible, as he had not looked at him very closely. He had not noticed if the boy had entered the flat during the short time he was himself in the kitchen. When the boy had gone, S. finished his coffee and then went back to the shop to clean the windows there. In doing so, he followed the same way back through the cellar.

From the relatively few bits of spontaneous information given by Lars in the interview with the child psychiatrist, it emerges that the boy, after visiting S.'s flat, went down into the courtyard where he met another boy. He would not give this boy's name, and when the psychiatrist asked what he had done with him, Lars answered: "I was not with him." After that, Lars went to the sweetshop round the block to spend the coin he claimed to have got from the window cleaner. When he arrived at the shop "the man" was already there cleaning windows.

On the occasion of a second police visit to the scene of the crime, a number of questions were put to Lars. The answers seemed to show that the boy had a good idea of what S.'s flat looked like. When asked if he had been inside the flat, he answered that he had been in the kitchen. Asked if he had been into the bathroom, he said that he had, and when asked what it looked like, he said there was a bath there. He added that there was a plug in the washbasin that you could pull and that this plug was not one of those with a chain. As the police confirmed that one could not see into the bathroom from the place by the front door where S. maintained the boy had stood, and that the construction of the plug, which the boy had described quite correctly, could not possibly be observed except from inside the bathroom, S.'s statement seemed to be false. S. was questioned further on this point, but he maintained that he had not noticed whether the boy had been into the bathroom on his own. S. denied as completely absurd the idea that he should have taken the boy into the bathroom and there committed the indecent act.

The "reliability test by the psychiatrist," based on the tape-recorded interview with the boy and the conversation with his mother, resulted in a categorical opinion in favor of the boy's reliability. In the words of the expert's summary:

The psychiatric investigation has shown Lars' statements to be reliable. He is bright and lively and of

normally developed intelligence. In spite of his mother's statement about the child's imaginativeness, he has shown no tendency to make up stories. Nor is he in any way maliciously disposed towards certain people, and his statements are consistent. There is no doubt that the man in question stuck his penis into the boy's mouth and that the boy sucked it. . . . I therefore certify that Lars' statements, that is, what his mother reported to the police and what came out of my investigations, are highly reliable.

The security investigation of S. was entirely in his favor. He had been liked in all the places where he had been employed, and his testimonials were exclusively good ones. The Russian connection came from the fact that he was born in a part of Finland that had been ceded to Russia at the last peace treaty. S. spoke good Swedish and had been engaged for 3½ years to a Swedish girl, whom he proposed to marry as soon as his naturalization had gone through. His fiancée and her mother were convinced that he had not committed the act of which he was accused. This was the view of all those who had had anything to do with S. since he arrived in Sweden.

In spite of the good testimonials, S.'s case could only be described as particularly bad at the conclusion of the police enquiry. He had indeed denied the indecent act, but he had admitted that a boy, who might be Lars, had visited him at the time in question. He denied that he had taken the boy with him into the flat, but the enquiry showed that the boy must have passed the same way as S. into his flat, and it was impossible for a boy of Lars' age to do this on his own. S.'s statement that the boy had only stood by the door was contradicted by the boy's ability to describe details in the bathroom, which he could not have seen from the front door. This evidence was further strengthened by the fact that Lars' own house had no such conveniences as were in S.'s. It could therefore be assumed that Lars had had no previous opportunity of seeing a washbasin plug of the type he had described. Finally, a test of the child's reliability had come to the conclusion that the indecency had taken place in the way the mother had said to the police.

INVESTIGATION FROM AN INTERACTION APPROACH

In this situation, the prosecution asked for a new specialist enquiry, feeling that the case needed further elucidation. The defendant's counsel had for his part not asked for any further enquiry.

The new enquiry was conducted by the author on different principles. Instead of focusing on the personality of the child witness, he followed an approach emphasizing analysis of social interaction in elucidating the circumstances and factors of personality that had combined to produce the charge against S.

The Mother

The first phase of the enquiry was a searching interview with the boy's mother, from which the following facts emerged:

Both parents considered that Lars was a very capable boy, advanced for his age, who took a lively interest in everything that happened around him. He had learned early to tell the truth, and, as far as his mother could judge, never lied, either to his father or mother. Besides, they had strictly forbidden him to lie. Their interest in the boy, until some two months ago their only child, was unusually strong. To show this, the mother told how the children in their district used to go round the houses and collect newspapers, which they later sold. The mother had, however, forbidden Lars to take part in such occupations unless an adult was with him. She had on a few occasions gone with him herself when he collected papers, and had then been careful to see that he bought nothing unwise with the money he earned. Both parents were very careful that he should not acquire any bad habits where money was concerned. As an example of this, the mother told how there was a dirty, badly cared for boy in the neighbourhood called Sune, who previously had played a lot with Lars. The mother had really been sorry for this Sune, who was universally unpopular in the district. Some time ago, however, she had forbidden Lars to be in Sune's company. The cause of this ban, which also included a ban on playing where Sune lived, was that she had once met the two boys together when Sune had had a whole krona (dollar). On this occasion, both boys had been lavishly provided with sweets, and when the mother questioned Sune, who was two years older than Lars, she gathered that Sune had filched the money. On one or more occasions before the incident with the window cleaner, Lars had been to Sune's house and played there, but his mother was convinced that he had not played with Sune since her prohibition. When she was asked if she knew whether Lars had ever collected newspapers on his own, she recalled that he had on one or two occasions been out and collected them together with older children. His mother did not know which houses they had visited on these occasions. When in this connection the investigator remarked that Lars was evidently quite experienced in going around ringing at the doors of strange houses to beg for newspapers, the mother conceded this without further discussion. She repeated, however, that she had forbidden him to do this by himself and that both she and the father were very particular that he obeyed their instructions.

As a further example of how alert and fearless her boy was, the mother said that Lars was very much

interested in different jobs and that he often played at doing such jobs as he could observe outside. Once in town he had gone up to a policeman and asked where "Mister Policeman" had put his radio-car. His mother said that he often talked to workmen about their work.

The major part of the conversation was devoted to the circumstances in which Lars' mother had talked to him on her return from shopping. What had specially surprised her was that the boy had shown no signs of discomfort or disgust. She had rather had the impression that he wanted to stay out and play in the street, when she and her mother showed signs of disquiet and took him indoors. His mother repeatedly reproduced the boy's tone and gestures when he came out with his sensational charge. Her impression was that Lars felt he was telling them something very interesting but that he did not at all understand the significance of this action. The correct version of his words was: "Just think, mummy, I had to suck his doodle."

The mother was also surprised that the boy could not be induced to repeat the accusation later. When his father came home he had tried to talk to the boy, who would, however, give no information about what the man had done to him. Even when his mother took Lars along to the police station in the evening to make her deposition, he would not tell anything to the policewoman who interviewed him. The interview had therefore been discontinued, but next afternoon a doctor (the psychiatrist) interviewed Lars. His mother then told the doctor all she knew, and, as she had understood it, the boy then told the doctor the whole story.

Regarding the boy's open fly buttons, the mother said that both women had been excited and bewildered at the discovery. Lars' mother could not recall whether her mother had framed her question to the boy in a leading way. She knew, however, that the grandmother was already suspicious of S. before the boy was questioned.

The Police

Of all the mother's disclosures, the most interesting to the investigator was that Lars had been interviewed by the police the very same evening the deposition was made. No notes of this interview were to be found in the police report. When the Inspector concerned was asked about it, it turned out that the interview had been considered a failure because the boy could not be induced to say anything. The notes had been thrown away, and, for various reasons that could not be more exactly specified, it had not been thought necessary to mention the "unsuccessful interview" in the report. After a special request, the policewoman concerned gave an account of what had happened. In addition to what was known previously, it transpired from her account that, in talking about the flat, the boy "had been there before and got newspapers, but at that time the 'lady' was there. Questioned about the indecent assault he became noticeably irritated and the earlier contact disappeared. . . . Only at the beginning of the talk, when Lars was specially interested in the typewriter, did he sit still on a chair. Later he ran around the room and played with the office furniture."

There was no doubt that a good deal of significant data lay in this discarded interview. Points which specially emerge are the boy's habits in collecting papers and his reluctance to repeat the charge itself.

The discovery that the police report was incomplete on an important point led to further investigations concerning the information it contained. Special interest was directed at the reproduction in direct speech of Lars' charges, which prefaced the report and made S.'s case more serious. Neither the choice of words nor the content agreed at all with the description given by the mother. Further questions revealed that this piece was a summary of the impression of the officials concerned at the time of the actual deposition, not a factual statement of what Lars had actually said.

According to the police report, the boy had said of his experience, "It was so disgusting." As appears from the conversation with the mother, the boy had never uttered such words but on the contrary gave the impression that he did not react emotionally. The form of words must therefore have originated from the author of the police report, who unintentionally amplified his account in accordance with his own expectations. This well-known phenomenon is chiefly notable because the amplification in this case has crept into the report in a form exceedingly dangerous for the accused. The conclusions of the test of the boy's reliability rested among other things on this statement of his feelings of disgust.

Lars: The Accuser

The next stage in the work of investigation was to study Lars himself. His degree of intellectual development was assessed by an expert in intelligence testing, with results summarized as follows:

At an intelligence test carried out on Saturday, June 6th, Lars scored an IQ of 123 (Terman-Merrill scale modified by Sandels). The experiment shows that it is specially easy to establish personal contact with Lars,

who gives the impression of being a child with healthy reactions. His curiosity is very great and his manner is spontaneous and marked by a certain easy self-assurance.

Lars was further examined by an expert in problems of suggestibility. (The results of this inquiry, undertaken in connection with a scrutiny of the incriminating tape-recorded interview, will be discussed later.) The author took part as an observer in some of the "suggestibility tests" and later amplified these observations with his own examination of the boy. It was easy to verify that the boy could only with difficulty be kept concentrating on the tasks he was given. He constantly slipped off his chair and roamed unhindered round the laboratory, examining the apparatus that stood within reach. Instead of answering the questions put to him, he asked his own questions about objects in the room and showed a pattern typical of many children of evasive tactics and distraction devices when in one way or another he felt unsure of himself.

The most interesting reaction came when the author during a pause in the suggestibility tests threw in the question: "Listen, Lars, do you know anyone called Sune?" Lars did not answer the question verbally. Instead he slipped off the sofa, went up to the author and gave him a crack on the jaw. It was not a real blow, but it was intended to represent a real crack on the jaw and fulfilled the function both of distraction and of symbolic demonstration of his antipathy to that particular question. Some five minutes later the question was repeated, whereupon Lars went through exactly the same performance. Since the situation seemed to be full of potential information, the question was repeated a third time but not until some ten minutes later, while other things were being dealt with. This time, the boy reacted by slipping off the sofa, so that he became invisible behind a table covered by a cloth. Down on the floor he said, drawing out the vowels: "Hoow noooosy you are."

A few days later, the author made a further personal examination of Lars. On this occasion as well, he demonstrated the distraction technique just described as soon as anything embarrassing came into the conversation.

After the investigation of Lars, the enquiry took a turn that had not been originally foreseen. So far, the author had not gone more closely into the question of S.'s possible innocence but had rather worked with the intention of studying the factual accuracy of the details of the charge. Such an aim was moreover supported by the police enquiry so very unfavorable to S.

A preliminary collation of the data so far produced by the psychological enquiry led, however, to trying out a completely new hypothesis. What suggested this idea more than anything was the boy's tendency to divert the adult's attention from whatever was embarassing him. A comparison of this tendency with the parents' rather naive idea that their boy never lied, and their overambitious accumulation of bans about his games and occupations, led one involuntarily to think of the possibility that Lars had simply used the assertion about the man's penis to divert his mother's attention from the things he then had on his conscience (he had visited forbidden territory, he had collected newspapers by himself, he had probably met Sune and perhaps also accepted money from him, he carried evidence of his crime under his arm and smeared round his mouth). Such an interpretation was further supported by his refusal to repeat the assertion that had caused so much trouble and excitement around him. The little boy, with all his wits about him, must soon have realized that his invention risked being revealed as a lie, and as lies were included in the list of forbidden things, he then made diligent use of his diversionary technique to avoid repeating the assertion.

S.: The Accused

The next task was to examine S. To begin with, S.'s employer, Andersson, was interviewed for information about S.'s conduct in connection with his arrest. In addition, the author obtained information from the police officers who took S. into custody, as well as from the Inspector who conducted the police enquiry. Then S. was brought from prison. He spent two days at the psychological laboratory, where his personality was examined. His intelligence was tested in the usual way and his life history reviewed in detail by two different specialists. His ability to reconstruct previous events was tested, and in a lengthy cross-examination his information concerning what happened on the fateful day and the day

after was reviewed. The investigation of S. also included a visit to his home where he was joined by his fiancée. In summary, S. showed himself to be a normally intelligent man, kind and decent to those around him, careful about himself and very exacting about how he appeared in the eyes of others. As he was waiting for naturalization, he was particularly anxious to avoid trouble and disputes. He was industrious, reliable, and ambitious in his work. His future was well planned. He and his fiancée had jointly saved up to be able to make a home as soon as he became a Swedish citizen. For the present, they lived with the fiancée's mother in her flat.

As far as the alleged crime was concerned, S. appeared completely unable to understand it. He constantly came back to the charge without prompting and lamented with curses and tears the absurd and humiliating aspect of his situation, but in between pulled himself together and hoped that the truth would help him. He made the utmost efforts to reconstruct what he had done the day of Lars' visit, and stuck obstinately even to such details as pointed suspicion against himself. (He refused, for instance, to agree that the boy could have been into the bathroom for a moment when S. in the kitchen had his back to the hall, and similarly he maintained that he had not heard anyone come after him through the cellar passage.) All his statements which could be checked by adult witnesses were shown to be correct.

S.'s sexual experiences were normal, and his way of talking about such things was that of the normal person. He showed no signs of homosexual thinking. On the whole, he was not interested in children. He was faithful to his fiancée and fond of the people close to him.

To the author he appeared incapable of dissimulating, giving the impression of a naively straightforward and sincere person, hurt and uncomprehending in face of the charge he had incurred. He had no reproaches to make against anyone but tried to understand the reactions of those around.

Regarding his demeanor when arrested by the police, the facts were these:

> The police went to look for him at midday at his workplace, but the firm's proprietor, Andersson, could not say exactly where S. was, as his cleaning programme included at least 30-odd firms in different places in the town. The police then informed Andersson that S. was wanted for a serious offense and that it meant several years prison for anyone helping the criminal to keep out of the way. Andersson, who took offense at this, suggested that the police should come round to the firm about five P. M., when S. usually came in after finishing work. When S. arrived at the firm's office about four P. M., he was met by the information from his employer that the police wanted him for a serious offense. S. was at first more surprised than anything else but then dismissed the matter, saying that there must be some mistake. Andersson, still excited by his conversation with the police, wanted S. to confide in him, so that the protection of the firm could be arranged in time. S. however calmed him down and assured him that he had not done anything and that a mistake had certainly been made. To put an end to the uncertainty he rang up the police on his own initiative, told them his name and asked for information as to why he was wanted. He managed to learn nothing from this conversation, as attested partly by Andersson and partly by someone who knew both S. and Andersson. S. then proposed that they should sit down and have a beer while they waited for the police. About half an hour later the police arrived and arrested S. At this there was some trouble, caused by Andersson's disapproval of the way the police behaved. S. himself made no difficulty. According to the Inspector, he had some outbursts later and behaved in an unbalanced way. S. himself said that when he came to the police station he had asked for permission to shave, as he was dirty after the day's work. He had then been given a shove and told to hold his tongue. "We decide things round here, not you," someone had said to him.

Resolution of Discrepancies

The investigation of S. left the hypothesis of his innocence still more promising than when the investigation of the boy was concluded. Nevertheless, this hypothesis could still not be accepted because of several important circumstances that remained unexplained. These were: (*a*) that Lars had been able to get through the locked cellar by himself; (*b*) that Lars could describe what the washbasin plug in S.'s bathroom looked like without having been in there; and (*c*) that Lars repeated his charge at the interview with the psychiatrist the day after the incident.

When the enquiry was resumed it was now directed at the unexplained details in Lars' experiences. Together with S. and the policeman who first investigated the scene of the crime, the author visited the house where S. lived. On this occasion, S. indicated the door through which he had passed into the cellar. This door was locked. It was not, however, identical with the door indicated by Lars. S. pointed out that the door he used to go through was the most natural one to choose

because it made direct contact with the staircase. From it a straight passage led to another door, which connected immediately with the staircase on which S. lived.

At the same time it was discovered that the door which the boy had previously indicated also led into the cellar and that via this door one could reach another door, which led out into the courtyard at the side of S.'s staircase. To the policeman's surprise it turned out that both these doors were unlocked on the occasion of this unannounced visit. A visit was therefore paid to the landlord's agent in the building. Asked about the closing of the cellar doors he answered: "Yes, they should always be locked."—"But they hardly ever are," added his wife, who stood beside him and evidently was less concerned with professional prestige than her husband.

After this breach in his defenses, information poured from the caretaker. He had a lot of trouble with the children in the house and children from the neighborhood, because they often played down in the cellar and carried on as they pleased. It was impossible to keep locked the doors which the children used, since people took their bicycles in that way and often forgot to lock up behind them. The agent had had particular trouble from a little hooligan called Sune Olsson. Once he had lit a fire in the cellar and another time he had damaged every bicycle that was down there. However, Sune had moved about three weeks ago, as they had managed to evict the family, who had made things very difficult, and who kept up a permanent feud with the neighbours and used foul and vulgar language.

To get an insight into Sune's living conditions, the author paid a visit to the house where he had lived. The new tenants said the flat had been badly looked after and that the previous tenants had gone out in a rush without anyone knowing where they had gone. The interesting thing about the flat was that, apart from being the other way round, it looked exactly like the one S. lived in. The bathroom looked precisely like the one the police had studied in S.'s flat, and the washbasin plug was the same. The author had no difficulty in imagining how the inquisitive little boy had run around in the Olssons' flat. The fact that he had been able to describe the plug to the police no longer seemed so extraordinary, and the association existing from the first between Sune's and S.'s homes also became more significant when the similarity became clear, because the distance between the two staircases was so great that an adult would hardly think of saying that they both lived in the same house.

Keen work by the police enabled them eventually to trace the Olsson family to a place a long way outside the town, where they had withdrawn from the demands of society and neighbors. Sune was then brought back to the family's old home, and asked to show the cellar he formerly played in. As soon as he got out of the car he went straight to the cellar entrance through which Lars claimed to have entered. In this cellar, where Sune evidently felt at home, the author talked to him about the playmates he had had. He knew at once who Lars was, and was able to describe him quite well. The last time they had met was inside in the courtyard, when Lars had some newspapers with him, said Sune. He supported this assertion by adding that he had left immediately afterwards. Asked if he had given Lars any money on that occasion, Sune gave no definite answer. It proved on the whole hard to get him to say anything if he suspected criticism behind the question. He behaved as children do who are accustomed to get the blame for everything that goes wrong: afraid, suspicious of strangers, at odds with society outside his own family.

Later the same day Sune's IQ was determined to be 92, quite sufficient for him to be able to understand the questions and answers exchanged with him.

Before Sune was taken back to his family's new retreat, his vocabulary was tested. It proved to be quite poor, but he had good knowledge of a number of vulgar expressions. He said he did not know the meaning of the word "doodle," but he knew several other names for the male sex organ. A number of hair-raising tales of how the mother had spoken to neighbors, recounted earlier by the agent, fill out the picture of the Olsson family's speech habits. There is therefore nothing against the assumption that Sune could have been the source of Lars' inspiration for the statement about how the window cleaner had acted. The fact that Lars in his version used the word "doodle" does not constitute a serious objection, because it must be assumed that Lars was intelligent enough to use such terms

in speaking to his mother as were sanctioned in the family's ordinary speech.

The question of where the coin had come from was still not definitely solved. According to Lars' mother, the agent and the candy-store clerks, it often happened that Sune had money. Logically, there is therefore hardly any need to adhere to Lars' statement that he had got the money from the "window-cleaner man." He could equally well have got it from Sune.

A few more small details remained which told against S. Such a detail was the unbuttoned trousers. But it is more usual for small boys to go around with their fly buttons undone after urinating on their own than for them to do them up again. For Lars in the present situation to blame the window cleaner was quite in accordance with his other behavior.

But wasn't it odd that Lars had rung just at S.'s door? Hardly, since at the time of day when this happened, most flats in the building were likely to have been empty. The special thing about S.'s door was that there was someone behind it who could open it and satisfy the little boy's wishes.

THE INCRIMINATING INTERVIEW RE-EXAMINED

The purpose of giving such a detailed account of a single case has been to illustrate the efficacy of two different working methods. While one case does not permit a conclusive assessment of their relative merit, the present study seems in the author's opinion to indicate that the description of events given by social interaction analysis is much superior to the personality analysis method of investigating witnesses.

In this connection, we may look more closely at the tape-recorded interview by the psychiatrist, as well as because it was here that Lars repeated his charge against S. The technical deficiencies in interviewing are to be understood in the context of the inexperience of the psychiatric expert in witness psychology, and the fact that the training of a child psychiatrist in Sweden is hardly directed at this type of task.

The analysis of the interview proceeded in two phases. To begin with, Lars' reactions to the kinds of questions used in the interview were explored in the "suggestibility test" previously mentioned. Then the questions and answers of the interview were reviewed critically in order to bring out what may have lain behind the boy's answers.

The "suggestibility test" was intended to assess (*a*) the tendency of the witness to state and sometimes even to experience sense perceptions that have no objective basis; (*b*) his tendency to yield to authorities enjoying a certain prestige; and (*c*) his tendency to answer certain types of suggestion with information that the interviewer is expecting.

The first type of test showed an obvious tendency on the boy's part to accept uncritically different kinds of sense perception. The record says, "One can plainly observe how the subject was induced to state auditory perceptions after quite short intervals, although there was no objective basis for such perceptions." The same phenomenon was established with tactile perceptions.

With prestige suggestions he showed less interest. The record says, "Lars was not interested and very inattentive in this test. He cannot be considered to have received any very deep impression from the interviewer's suggestions. When he accepted the interviewer's proposals, it seemed almost to be in order to get away from the whole thing as quickly as possible."

The test with different kinds of suggestive questions was of the greatest importance for the enquiry. The questions were partly anticipatory of the type, "You do see . . ." or "Didn't you see that . . .," partly alternative questions of the type, "Was the car blue or green?" when the car was in fact red. Finally, simple *yes* and *no* questions were tested, which also have a certain suggestive effect, as they are more often answered in the affirmative than the negative.

The record shows that Lars could be brought to assent to statements that had no basis in his own observations. He could even be induced to demonstrate with sounds and gestures things which did not exist. For example, the interviewer asked Lars if he had not seen a small pair of scissors on the experimental table. Lars said he had, although there were no scissors among the objects he had seen. When the interviewer asked the boy how big they were, Lars demonstrated the approximate size of a pair of scissors with his hands. The interviewer then asked if they were not bigger

and measured out a bigger size with his hands. The willing little boy then himself measured out a bigger size with his hands and answered: "Yes, like this."

The examination of the tape-recorded interview shows that the questions put were in the main anticipatory or framed as statements to which the boy had only to say *yes* or *no*. Instead of sorting out what really had happened, the interviewer tried to get the boy to repeat his charge. The result of the interview, to summarize, was a repetition of the mother's statements put into the child's mouth. The accusation only came up when the interviewer had used both bribes and pressure.

Dr.: But it would be nice to get straight just what happened with that man. Then you can go home with your mummy and take the aeroplane with you, and we won't have to meet again.
Lars: Whose are these toys?
Dr.: The toys are mine, and we've promised you that you can take an aeroplane with you. But you must tell just exactly what happened with the man, or perhaps you want another of the toys, but of course you would rather have the aeroplane, wouldn't you?

After the dialogue just quoted, Lars managed to distract the interviewer's attention, and the talk ranged for a long time over peripheral subjects. In the end the boy wanted to go to his mother.

Dr.: But first you can surely tell me something about the man. Then you can go home with your mummy.

This attempt succeeded no better, however. After a break, when the boy talked to his mother for a while, the interview was resumed, and now the interviewer altered the situation. He proposed in fact that Lars should tell him what he had previously told his mother.

Dr.: Now what was it you told your mummy? When mummy met you? Can't you tell me about it too?

If one accepts that the boy lied the day before and was now afraid to lie once more, the interviewer's request must have seemed inviting, for to talk about what he told his mother the day before hardly seems like a new lie from a child's point of view. In spite of this it took no less than four different requests from the interviewer before the boy, guided by the interviewer's yes and no questions, began to take up the assault story. It is striking to see how he is able to dominate the adult even in this situation. When Lars answered the question: "Did you go into any room with the man?" with "Yes, I went into their bathroom," the interviewer became more careful and avoided putting answers into the child's mouth. When the boy was thus left with no pointers as to what answers the "doctor man" wanted, he tricked the interviewer back into the old technique.

Dr.: Did the man do anything with you or did you do anything with the man, or what did happen in the WC?
Lars: *Guess!*

It is hardly surprising that after this invitation the interviewer managed to guess what Lars had told his mother the day before. The central element in the interview was thus a reproduction of what the mother told the police and the first interviewer.

On the question of the boy's first acquaintance with S. the interview contains a number of statements which depend on the fact that the boy had not been addressed by S. but had followed him on his own initiative. The following exchange shows this clearly:

Dr.: So you stayed both up there on the stairs?
Lars: (silence).
Dr.: Well?
Lars: No, I went in to him.

Another exchange in the same vein is the following:

Dr.: Yes, so the man came? [that is, up to the boy and asked him into the flat in accordance with the interviewer's presupposition].
Lars: *He was there.*

The analysis of the interview clearly shows that the conclusions drawn by the interviewer go far beyond what the material legitimately supports. An experienced and skillful interviewer would never have drawn his conclusions with such assurance and would probably have produced several leads calling for further investigation; it nevertheless appears improbable that a real solution of the problem could have been achieved without an extension of the enquiry to the plane of social interaction.

WHAT HAPPENED: A RECONSTRUCTION

A man is cleaning a shop window in a suburban street. A little way off a small boy stands watching him. He would like to be a

window cleaner too. The man finishes work. He collects his things and his bicycle but is stopped in the street by another man who begs for money. The little boy goes up and asks the man if he may carry his ladder. The boy often does this sort of thing. He talks to policemen and workmen. "No," says the man, "it is too heavy for you."

He is a nice man, he does not snap at anyone, children or beggars. Then he pushes his bicycle round the corner and past the next block, as far as the sweetshop, right by the house where he lives. The little boy wanders after him, as curious as ever. He sees the man put his bicycle and ladder against the wall of the shop. Then the man disappears into a house further off, where the boy used to play before mummy told him he must not go so far away.

The man has disappeared, but the boy goes into the house too. He goes through the cellar door, where children often do go in. He goes on through to the other side and out through the bicycle exit and comes into the courtyard. He has been there before too. He has been collecting newspapers in this house. His mother is shopping, and there are a lot of things that can be done without her knowing.

He goes up into the house and rings at the doors, but nobody opens. Not until he comes to the third floor is there a door that opens, and the man who opens it is the one who cleans windows. The boy asks for papers, and the nice man goes into the kitchen to fetch some. The little boy sees that this is just like where Sune lives, further off in the building. Two rooms and kitchen, with a hall and bathroom. He knows that is what Sune has, because he has been home to Sune's and played in his apartment, before mummy said he must not play with Sune. Then the man comes with the papers and gives them to him. He goes down the stairs with them under his arm. At first his thoughts are full of the window cleaner and that he has been to the man's apartment. But then he thinks that that is because he has gone with him into the house and gone up to his apartment.

When the boy comes out into the courtyard, Sune is standing there. And Sune asks where he got the papers. But when the boy tells Sune that he got them from a man, Sune tells him something old men do with children when you go to their place. And the little boy thinks that he has been to a man who does that.

Then he gets a big black penny from Sune, which Sune found at home, and he leaves Sune and goes around the house, which is ever so long, and when he comes around the other corner, he sees the man standing on his ladder cleaning the sweetshop windows.

Then he goes into the shop and spends the penny on a bit of liquorice. It buys him no more than half a bit, because the lady says there is not enough for more. He looks at the man who gave him the papers, and when he goes out of the shop he thinks about what Sune told him about old men, and then he goes home. He is not thinking of anything much any more, but he has a feeling in his stomach that he has done something he should not. Collecting papers. Being with Sune. And what Sune told him about. And then he meets mummy and granny, and they ask about the papers he got from the window-cleaner man.

"But how do you know him?" says mother anxiously.

And the little boy explains.

"I went home with him," he says.

If a man asks you in with him to collect something, no one can help that.

"And the candy," says mother, "where did you get that?"

He is not allowed to be with Sune. Perhaps it was the window cleaner who gave him the money. Old men give little children money. You go home with them, and then you get a penny.

"The window cleaner," he says.

But mother and grandmother are getting upset and hushing him up and pulling him to the stairs. There is something wrong with what he said. But he does not want to go up now, and so he gives his mother his sweet look and says:

"Just think, mummy, I had to suck his doodle."

It is a smack on the jaw for mother and grandmother, the biggest smack in the world.

But he still does not manage to stay out. Grandmother sees that his fly buttons are undone. But the window cleaner did that. Mummy talks to daddy on the telephone. The police will take the window cleaner, and he will say that Lars did not suck his doodle at

all. So the boy says nothing to his daddy, when daddy asks after dinner. When he goes with mummy to the police, and the lady who has a typewriter talks to him, he does not tell her anything either.

But the police and the other grown-ups are catching the foreigner who has committed an indecent assault. In the end, a doctor comes and helps the police. He delivers his judgment after being two hours with the little boy, who besides being bright and inquisitive is not afraid of the silly grown-ups who only understand bits of his thoughts.

And society prepares to put out of harm's way someone who opened a door and gave a boy a dozen papers. It is very interesting and at the same time somehow thrilling. As drama, it is oppressive and horrifying.

It must be added that the complete reliability investigation led the prosecutor to decide that there was no case against S. After six weeks, S. left prison a free man. When he was set free no one told him why he was not to be kept in jail any longer.

Received June 10, 1957.

CRITIQUE AND NOTES

ACCURACY OF SELF-PERCEPTION AND GROUP-PERCEPTION AS RELATED TO TWO LEADERSHIP ROLES[1]

DAPHNE E. BUGENTAL AND GEORGE F. J. LEHNER

University of California, Los Angeles

In the extensive literature on leadership, there is a frequent tendency to treat leadership as a unitary concept that may be measured by a wide variety of sociometric questions, ratings by peers or supervisors, observational techniques, etc. Consideration is seldom given to the possibility that different types of leadership or status are measured when different criteria for leadership are used. Gibb (4), an exception, found that persons who were socio-telic or psyche-telic choices of other group members were different from individuals found to be leaders by observational techniques. He also found that individuals viewed by the group as having "influence" more closely compared with the leaders selected by observational techniques than with the socio-telic or psyche-telic choices. Bales (1) also gave consideration to the possibility of different leadership roles. In observing groups over time, he found that the role of "best-liked" group member became quite distinct from the "leader" role.

Two types of status or leadership roles appear to be distinguishable. The first is the role of the true "leader" whose primary function is the guidance of the group towards their goals. The second "leadership" role is that of the popular group member, which is roughly equivalent to Bales' "best-liked" role and Gibb's "psyche-telic" and "socio-telic" choices. The popular person is the center of positive affect within the group. However, hostility may be directed toward the individual enacting the "leader" role. The popular person does not assume responsibility for reaching group goals, and he interacts on an emotional or "social" level rather than the task-oriented level of the leader. These two roles may overlap and the same person play both roles, but they are distinct roles. Probably they are more dichotomous in largely task-oriented groups.

In view of the characteristics of these two roles, the popular member would appear to be in the better position to perceive accurately other group members and the group as a whole; that is, he should be more likely to know how well other group members are liked or how they feel about significant group issues. His greater "insight" is predicted from the sociocentrality of his role within the group.

Two studies concerned with the ability of leaders to estimate group opinion found discrepant findings. Chowdhry and Newcomb (2) found that "leaders" (defined as individuals sociometrically chosen on questions measuring both popularity and leadership variables) had better insight than isolates or nonleaders on group-relevant issues. Hites and Campbell (7), on the other hand, found that "leaders" (defined as elected or appointed leaders) did not differ significantly from nonleaders in ability to estimate group opinion on group-relevant issues. These differences in findings, to the extent that the reported findings are free of artifactual components, may have been influenced by the two ways of measuring leadership.

Several studies (3, 5, 8, 9) have been concerned with the insight of leaders (most frequently measured by popularity variables) into self and others on social and personality variables. In general, leaders were found to have better insight than nonleaders into others. The evidence for greater self-insight was not strong, although there was some evidence for greater self-insight with higher status.

The intended contribution of the present study is the prediction that it is the popular person, not the leader, who has more accurate perception of the group structure and of group members. It is also predicted that the popular person has more accurate self-perception than does the leader. Our main interest in testing this hypothesis is to determine whether the popular group member is generally more "insightful" or if his accuracy of perception is related specifically to his role within the group. If popular individuals have both more accurate self-perception and more accurate group-perception than do leaders, they would appear to be more "insightful" in general. If they have more accurate perception of the group but not of themselves, their "insight" would be seen as primarily associated with their position within the group. Our specific hypotheses are:

Hypothesis 1. Popular group members are more accurate than leaders in their perception of status relationships within the group.

Hypothesis 2. Popular group members are more accurate than leaders in their perception of significant group dimensions.

[1] Appreciation is expressed to R. T. Morris for his suggestions in the conducting of the study and for the use of his students as *Ss*.

Hypothesis 3. Popular group members have more accurate self-perception than leaders on personality variables.

METHOD

Members of a coeducational upper division sociology class served as subjects (*Ss*). The class ($N = 48$) was randomly divided into five groups of 8 to 11 members each with the purpose of completing a class project. These groups had been meeting for a semester at the time of testing.

In order to distinguish between "leaders" and "popular" group members, we drew on Bales' (1) discussion of leader roles to devise a questionnaire consisting of a series of descriptions of members, part of which related to the "leader" role and part to the "popularity" role. Examples of "leader" descriptions are:

1. His (her) suggestions for group action are usually accepted and followed by the group.
2. He (she) assumes responsibility for getting group work done and group problems solved.

Examples of "popularity" items are:

1. He (she) is well-liked by almost all, or all group members.
2. Group members seek him (her) out as a pleasant person to be around outside the group.

Each person in a group rated all other members of his group on eighteen items of this type on a five-point scale. *Ss* were instructed to check "A" if the description was "definitely true" of the member they were rating, "B" if it was "mostly true," "C" if it was to an equal degree true and false or if they were unable to say, "D" if it was "mostly false," and "E" if it was "definitely false." Ratings were thus obtained on all *Ss* on both leadership and popularity variables.

All *Ss* were then ranked for mean rating received on leadership and on popularity. *Ss* in the top 25 per cent on leadership ratings were defined as "leaders"; those in the top 25 per cent on popularity ratings, as "populars." As we wished to distinguish between the "leader" role and the "popular" role, overlapping cases ($N = 6$) were eliminated. All further comparisons were based on seven leaders and five popular members.[2]

To determine whether popular group members were more accurate than leaders in their perception of group status relationships, we again used our group member descriptions questionnaire. The closeness of agreement between a member's descriptions and those of the group as a whole provided a measure of consensual accuracy, in terms of a mean error score based upon the differences between a person's ratings of other group members and the group's ratings of these same members. Insight into popularity and into leadership were measured separately.

To obtain measures of accuracy of perception of significant group dimensions, we used the Hemphill Group Dimensions Descriptions Questionnaire (6). The Hemphill questionnaire measures 13 group dimensions, such as viscidity, hedonic tone, autonomy, etc.; one of these dimensions (permeability) was omitted as inappropriate. The questionnaire requires members to rate their group on a five-point rating scale on a series of statements such as, "The group has well understood but unwritten rules concerning member conduct." Again we measured accuracy consensually, so that an individual who was very discrepant from the rest of the group in his ratings of the group had a high mean error score, and one who agreed closely with the group as a whole had a low mean error score.

Accuracy of self-perception on personality variables was measured with the Guilford-Zimmerman Temperament Survey and a self-rating test in which *Ss* rated themselves on the factors measured by the Guilford-Zimmerman. *T* scores for the leaders and populars based on distributions of self-ratings on each factor for the total group were compared with *T* scores received on the same factors as measured by the Guilford-Zimmerman, and difference or error scores were computed.

TABLE 1

MEDIAN ERROR SCORES[a] ON RATINGS OF GROUP MEMBERS AND GROUP DIMENSIONS

Variables Rated	"Populars" ($N = 5$)	"Leaders" ($N = 7$)	U
Popularity of others	.59	.87	7*
Leadership of others	.85	.77	16
Hemphill group dimensions	.56	.87	3**

[a] Error scores represent mean standard scores.
* Significant at the .05 level of confidence.
** Significant at the .01 level of confidence.

RESULTS

In comparing leaders with popular group members for accuracy of perception (indexed by error scores) we used a one-tailed nonparametric measure, the Mann-Whitney U test. In terms of our hypotheses, our results were as follows:

Hypothesis 1. Popular members, as predicted, were more accurate than leaders in their perception of the popularity of others. ($U = 7$, $p < .05$.) However, there was no difference between leaders and popular group members in accuracy of perception of the leadership status of others. ($U = 16$, n.s.)

Hypothesis 2. As predicted, popular members were more accurate than leaders in their perception of significant group dimensions. ($U = 3$, $p < .05$.)

Hypothesis 3. Popular group members did not differ from leaders in accuracy of perception of self on personality variables. ($U = 17$, n.s.)

DISCUSSION

In the light of these results it appears that individuals enacting the role of the popular member have more accurate perception of the popularity

[2] Two other populars had to be eliminated because of incomplete returns.

status of others and of significant group dimensions than do members enacting the leader role. It is difficult, however, to say whether the popular person achieves this role because of his more accurate perception of the group, or whether he has more accurate perception of the group because of his position within the group. Both relationships are plausible.

Apparently, the popularity role does not include more accurate perception of the leadership status of others. The well-liked person appears to have superior accuracy of perception only with respect to factors specifically related to his role. His understanding of the group relates to how well other members are liked, how members feel about the group, etc.—but does not include a superior understanding of who is directing the group towards their goals. Additional evidence for the belief that the popular person's more accurate perception is specifically tied to areas related to his role is given by our finding that there is no difference between the popular person and the leader in accuracy of self-perception. The "insight" of the popular person is not general, then, but is confined to variables associated with the sociocentrality of his role.

SUMMARY

"Popular" group members were compared with "leaders" of five small groups for accuracy of perception of other group members on popularity and leadership variables, accuracy of perception of the group as a whole on 12 group dimensions, and accuracy of perception of themselves on personality variables. "Popular" members were found to have significantly more accurate perception of others on popularity variables, and of the group as a whole on the group dimensions measured. No significant differences were found between "popular" members and "leaders" in accuracy of self-perception on personality variables or in accuracy of perception of others on leadership variables.

REFERENCES

1. BALES, R. F. The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shils (Eds.), *Working papers in the theory of action*. Glencoe, Ill.: Free Press, 1953. Pp. 111–161.
2. CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
3. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18.
4. GIBB, C. A. The sociometry of leadership in temporary groups. *Sociometry*, 1950, **13**, 226–243.
5. GREEN, G. H. Insight and group adjustment. *J. abnorm. soc. Psychol.*, 1948, **43**, 49–61.
6. HEMPHILL, J. K. *Group dimensions*. Columbus, Ohio: Bur. of Business Res., 1956.
7. HITES, R. W., & CAMPBELL, D. T. A test of the ability of fraternity leaders to estimate group opinion. *J. soc. Psychol.*, 1950, **32**, 95–100.
8. LEMANN, T. B., & SOLOMON, R. L. Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry*, 1952, **15**, 7–90.
9. NORMAN, R. D. The interrelationships among acceptance-rejection, self-other identity, insight into self, and realistic perception of others. *J. soc. Psychol.*, 1953, **37**, 205–235.

Received April 29, 1957.

BEHAVIOR AND GRAMMATICAL STYLE

LEONARD W. DOOB

Yale University

Do people express themselves in such characteristic ways that a relatively simple analysis of their grammatical style can yield useful clinical insights? Style as well as content are said to spring from enduring predispositions: "the frequency with which a person employs a certain grammatical category such as nouns, verbs, adjectives and the like has been shown to be related to various personality characteristics" (8, pp. 248–249). The evidence cited after that assertion and existing evidence in general are spotty. By an ex-post-facto analysis, for example, the behavior of two *S*s has been linked to their prose style (9). Or the grammar employed by various types of abnormal patients has been partially differentiated (1) or compared with normal *S*s (3). A connection has been sought between a single category like verb-adjective ratio and a broad behavioral category like emotional stability (2).

METHOD

To explore a wide variety of possible relations between behavior and style, *S*s whose behavior had already been experimentally and heterogeneously sampled by Linton (5) were used.[1] A total of 28 measures was obtained from 49 male students at Brooklyn College. The measures are grouped and listed in Table 1.

[1] Keen appreciation is expressed to Harriet B. Linton for placing the behavioral data at my disposal, for collecting the essays, and for numerous suggestions.

TABLE 1
BEHAVIORAL MEASURES

Perceptual Situations

1. (−) Ability to locate Embedded Figures
2. (−) Ability to respond accurately in Tilting-Room-Tilting-Chair Test*
3. (−) Factorial ability to handle object in field under distracting conditions*
4. (−) Factorial ability to handle field as a whole

F-Scale and Other Questionnaires

5. (+) Conformity to convention*
6. (+) Submissiveness to parents*
7. (+) Submissiveness to authoritative figures
8. (−) Recognized self-assertiveness*
9. (+) Disguised aggressiveness

Autokinetic Situation

10. (−) General ability to adapt to the situation
11. (+) Report of relatively great movement
12. (−) Relatively fast reaction time*
13. (+) Influenced by *E*'s confederate
14. (+) Efforts to counteract *E*'s confederate*
15. (+) Generally impressed by *E*'s confederate

Verbal Communication

16. (+) Change in advocated direction

Syllogisms

17. (−) Ability to reason correctly concerning neutral items*
18. (+) Susceptibility to "atmosphere" effect

Rorschach

19. (−) High Total R
20. (+) High Total T
21. (−) High Total M*
22. (+) High Hd:H
23. (+) High F%
24. (+) High Total C
25. (−) High M:C
26. (+) High (C + CF): FC
27. (−) High Total W and/or P
28. (−) High "mammals":"non-mammals" A

TABLE 2
PROSE CATEGORIES

1. (+) Number of words*
2. (−) Words per sentence
3. (−) Variability of sentence length
4. (−) "Noun-qualifiers" (adjectives, participles, dependent clauses) per 100 words
5. (−) "Noun-qualifiers" per sentence*
6. (−) "Adverbs" (excluding negatives) per 100 words*
7. (−) "Adverbs" per sentence
8. (−) "Adverbs" per 100 "noun-qualifiers"*
9. (−) "Pronouns" (excluding ego references, including adjectives) per 100 words*
10. (−) "Pronouns" per sentence*
11. (−) "Pronouns" per 100 "noun-qualifiers"*
12. (−) "Pronouns" per 100 verbs*
13. (+) "Ego pronouns" (including adjectives) per 100 words
14. (+) "Ego pronouns" per sentence
15. (+) "Ego pronouns" per 100 "pronouns"
16. (−) Verbs per 100 words
17. (−) Verbs per sentence
18. (+) Verbs per 100 "noun-qualifiers"
19. (+) Forms of "to be" per 100 "principal verbs" (carrying the sentence's principal action, usually of the independent clause)
20. (−) Forms of "to have" (not as auxiliary) per 100 "principal verbs"*
21. (−) "Principal cognitive verbs" (involving internal reactions) per 100 "principal verbs"*
22. (−) "Principal conative verbs" (involving overt action) per 100 "principal verbs"*
23. (+) "Principal verbs" in future tense per 100 "principal verbs"
24. (+) "Principal verbs" in passive mode per 100 "principal verbs"*
25. (+) Concrete or specific "principal subjects" (the subject of the sentence's "principal verb") per 100 "principal subjects"
26. (−) Negative sentences per 100 sentences
27. (−) Sentences with "thought qualification" (clauses, phrases, adverbs, subjunctive mode) per 100 sentences.

Details concerning the "perceptual situations" can be found in Witkin *et al.* (10), concerning the other situations in Linton (5, 6, 7). Many of the measures have been found by these investigators to be significantly interrelated.

Samples of the *S*s' prose style were essays in which they described their choice of a major subject in college and their proposed career. The prose was analyzed by means of 27 categories which are relatively devoid of content and which were chosen either because they could be easily applied to any communication or because they had been used in previous studies. They are listed in Table 2. Those defined slightly differently from conventional grammar are enclosed in quotation marks and, upon their first appearance in the table, they are accompanied by an operational definition in parentheses. The reliability of the categories—tested by sampling agreement between two independent coders—is high, modally in the .90's.

RESULTS

Linton's behavioral data have been dichotomized at psychologically meaningful points whenever possible. The cut-off points for the grammatical categories have been arbitrarily established to achieve approximately equal numbers in each division. The 28 behavioral measures have then been paired with the 27 grammatical categories. A careful inspection of the chi squares which have been calculated or estimated for the resulting 756 2 x 2 tables reveals that the number of significant differences at the .05 level with a two-tail test does not exceed chance expectancy and that the few significant differences do not form useful or meaningful patterns. More refined statistical analysis scarcely seems worthwhile.

One additional analysis, however, has been made by postulating enduring dispositions which may affect both behavior and grammatical style. A freely and loosely modified version of Witkin's and Linton's psychological dichotomy has been employed, but for convenience Witkin's concepts are borrowed: a relative tendency to be *passive* ("associated with field-dependent perceptual performance") versus a relative tendency to be *active* ("associated with independent or analytical perceptual performance") (10, p. 467). In Table 1,

the sign in parentheses before each measure provides a rough summary of Linton's findings: a plus sign indicates that those with passive rather than active tendencies demonstrated the behavior, a minus sign signifies the reverse.

In Table 2, the signs in parentheses suggest whether those with passive tendencies (+) or with active ones (−) can be expected to produce a *higher* frequency of the prose category. Almost all of these predictions have had to be based upon arbitrary assumptions and intuitions rather than previous research. "Passive" *S*s, it has been presumed, would cooperate more fully by writing more words [# 1] and, since the task involved in part their future career, by using the future tense more frequently [# 23]. They might be expected to refer to themselves fairly often [#'s 13–15], to be more interested in general action than in specifically changing the environment [#'s 18, 19], to abstract less [# 25], and of course to write in the passive rather than the active mode [# 24]. In contrast, "active" *S*s, it was supposed, would write more complicated prose as judged by sentence length [# 2], sentence variability [# 3], and the rather pronounced use of "pronouns" [#'s 9–12]. It might be anticipated, too, that they would seek to control their environment and themselves by modifying and qualifying the former [#'s 4–8, 26, 27]; and that they would use more verbs in general [#'s 16, 17] and especially those indicating activities of possessing, feeling or thinking, and overt movement [#'s 20–22].

Even though some of the measures are not independent, a simple sign test has been applied to the predictions in both tables. When the *S*s are divided into plus and minus groups on the basis a single category in one table, the plus group should contain a higher proportion of plus scores than the minus group from all or almost all the categories of the other table. The "ability to locate Embedded Figures," for example, tends to be possessed by "active" *S*s (Category 1, Table 1, is designated minus). Those *S*s demonstrating superior ability in this respect, therefore, should write fewer words since "number of words" is presumed to be associated with "passive" *S*s (Category 1, Table 2, is designated plus); they should write longer sentences (Category 2, Table 2, is minus); and so on, through the 27 categories of Table 2. Thus, each behavioral category can be viewed as being tested 27 times by means of the grammatical categories; and each grammatical category 28 times by means of the behavioral categories. Items in Tables 1 and 2 which are followed by an asterisk show significant trends at the .05 level, one-tail test. In Table 1 it can be seen that 9 of the behavioral measures show such a trend; in Table 2 the number of significant grammatical categories is 12.

DISCUSSION AND CONCLUSION

The results from applying the sign test to the grammatical categories can be briefly summarized. In contrast with those revealing a relative tendency to be passive in the behavioral situations the *S*s revealing a relative tendency to be active are inclined (*a*) to qualify statements by means of "adverbs" and "noun-qualifiers"; (*b*) to use "pronouns"; (*c*) to use verbs indicating action, possession, and internal reactions; and (*d*) to be comparatively less verbose. It is noteworthy that some parts of that summary contradict and others confirm previous studies. Here, for example, the verb-adjective quotient (Category 18, Table 2) is not discriminating, but the passive-mode index (Category 24) is; whereas apparently Balken and Masserman in their frequently cited study (1) obtained reverse results when they compared three classes of neurotics and psychotics. In both studies, however, number of words is a significant measure.

The failure of the present study to reveal more numerous significant relations per se might be variously explained. Faulty sampling may be involved: the *S*s may be too homogeneous, or Linton's behavioral measures or the pieces of prose may not have adequately reflected their enduring dispositions. In spite of the large number of indices, the grammar of the prose may not have been validly apprehended. Possibly scaling techniques are needed; those which have been here attempted—by various combinations of the grammatical categories—have not been fruitful.

Resorting to the sign test, nevertheless, produced trends which suggest that grammatical style has at least some tenuous connection with personality. That connection, however, is not specific enough to justify the statement of McClelland and his associates which was quoted above and which they themselves explicitly modify. For it varies with (*a*) the kind of behavior being observed or measured; (*b*) the categories being employed to analyze the grammatical style; and (*c*) as Goldman-Eisler (4) has shown for psychiatric patients being interviewed by different psychiatrists, the kind of situation in which the verbal habits are sampled. In addition, successful analyses of grammatical style cannot be made in the course of an interview or through any rapid device; the present categories, even though selected in part because they appear simple, require tedious counting and statistical analysis. For the moment, therefore, the following practical conclusion seems warranted: the analysis of gram-

matical style by itself is not a useful or feasible clinical instrument.

REFERENCES

1. BALKEN, E. R., & MASSERMAN, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neuroses. *J. Psychol.*, 1940, **10**, 75–86.
2. BUSEMANN, A. *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik.* Jena: Gustav Fischer, 1925.
3. FAIRBANKS, H. Studies in language behavior: II. The quantitative differentiation of samples of spoken language. *Psychol. Monogr.*, 1944, **56**, 19–38.
4. GOLDMAN-EISLER, F. A study of individual differences and of interaction in the behaviour of some aspects of language in interviews. *J. ment. Sci.*, 1954, **100**, 177–197.
5. LINTON, H. B. Relations between mode of perception and the tendency to conform. Unpublished doctoral dissertation, Yale Univer., 1952.
6. LINTON, H. B. Rorschach correlates of response to suggestion. *J. abnorm. soc. Psychol.*, 1954, **49**, 75–83.
7. LINTON, H. B. Autokinetic judgments as a measure of influence. *J. abnorm. soc. Psychol.*, 1954, **49**, 464–466.
8. MCCLELLAND, D. C., ATKINSON, J. W., *et al. The achievement motive.* New York: Appleton-Century-Crofts, 1953.
9. SANFORD, F. H. Speech and personality: a comparative case study. *Character & Pers.*, 1942, **10**, 169–198.
10. WITKIN, H. A., LEWIS, H. B., *et al. Personality through perception.* New York: Harper, 1954.

Received May 9, 1957.

ACHIEVEMENT STRIVING, SOCIAL CLASS, AND TEST ANXIETY[1]

MARTIN L. HOFFMAN
Merrill-Palmer School

SPIRO B. MITSOS
Evansville State Hospital

ROLAND E. PROTZ
Beatty Memorial Hospital

NEED Achievement scores of middle-class, as compared to working-class, adolescents have been found by Douvan (1) and Rosen (3) to be high under ordinary test conditions; and by Douvan (1) to remain at the same level even when a desired prize is offered for superior performance. Thus, achievement striving in the middle class appears to be highly internalized and relatively unaffected by external reward.

On the assumption that achievement striving relates positively to performance, it was hypothesized that middle-class performance would also be relatively unaffected by external reward; more specifically, that compared to working-class *Ss*, middle-class *Ss* would show little improvement in test performance under financial incentive. Two studies were done, one using a simple motor test (Study A) and the other, a written intelligence test (Study B).

METHOD

Social Class

The details of the class measure are presented by Douvan (1). In brief, *Ss* designated as working class were those who identified with the working class and whose fathers did manual work; those designated as middle class identified with the middle class and their fathers' occupations largely involved dealing with people.

Study A

The procedure of Study A followed that used by Douvan (1). Students in their regular high school classrooms performed a simple number-tracing task under a time limit that guaranteed that no one would finish. In five classes the test was introduced simply as a measure of "visual-motor coordination, how well your hands and eyes work together." For six other classes it was added that in order to encourage the students to try their hardest, a five-dollar reward would be offered to every one scoring above "a certain point." In order to minimize competitive motivation, the possibility of everyone's getting the reward was emphasized.

The *Ss* in Study A were high school juniors and seniors in a medium sized midwestern city (population about 80,000). From a pool of 271 students, 122 were selected who clearly fit the middle-class criteria and 37 who clearly fit the working-class criteria. It turned out that 68 middle-class and 24 working-class *Ss* had taken the test under the financial reward condition while 54 middle-class and 13 working-class *Ss* had performed under the no-reward condition.

Study B

In Study B, a before-after design was used. Alternate forms of the Otis Quick Scoring Mental Ability Test

[1] Part of this paper was read at the American Psychological Association meetings in 1957.

were given the same *Ss* under the two conditions. The setting in each case was a large assembly hall in which all students took the test together. Gamma Test Form C was administered first. The procedure given in the Manual of Directions was followed except that the students were allowed to work for only ten minutes. The instructions to the students stated that the test measured one's "ability to think." Gamma Test Form C was given one week later with no advance notice that there was going to be a retest. No information was provided regarding previous level of performance. The instructions included mention that the test was similar to the one taken previously, that we were now interested in how much they could improve their scores, and that five dollars would be given to "those of you who do a good deal better on the test than they did a week ago." To minimize competitive motivation, it was added that "all sizeable improvements would be rewarded."

A questionnaire asking each *S* to estimate (*a*) the extent of his desire for a five-dollar prize and (*b*) how hard he would work for this sum, was used to provide an index of conscious motivation aroused by the reward offer.

The Otis was scored in the standard manner, i.e., adding the number of items answered correctly. To take into account changes in the number of errors, the tests were also scored in terms of right minus wrong.

The *Ss* in Study B were high school seniors in a small midwestern city (population about 15,000), none of whom had ever taken the Otis before. From a pool of 150 students, 47 middle-class and 27 working-class *Ss* were selected using our dual criterion.

RESULTS AND DISCUSSION

Study A

The data for Study A (Table 1) confirm the hypothesis that the performance differential under the two test conditions would be smaller for the middle class than for the working class. The working class improved significantly in the reward condition while middle-class scores declined slightly.

Study B

In Study B, with either total-correct or right-minus-wrong scoring of the Otis, both the middle- and working-class *Ss* obtained significantly higher scores in the reward than in the non-reward condition (Table 2). The middle class showed less improvement, as expected, but the difference is not significant ($p < .3$ with total-correct scoring and $p < .2$ with right-minus-wrong scoring). The hypothesis of less improvement in the middle class receives statistical support, however, when we disregard the amount of improvement and consider only the percentage of *Ss* who show any improvement at all. Thus, with total-correct scoring, 85.2 per cent of the working-class and 61.7 per cent of the middle-class *Ss* showed an improvement ($p < .02$). With right-minus-wrong scoring the percentage of working- and middle-class *Ss* showing improvement are 77.8 and 55.3 respectively ($p < .03$).

TABLE 1

MIDDLE-CLASS AND WORKING-CLASS PERFORMANCE ON A MOTOR TASK UNDER TWO REWARD CONDITIONS

	Material Reward				
	Absent		Present		
Class	*N*	Mean	*N*	Mean	*p*
Middle	54	20.11	68	19.87	—
Working	13	16.00	24	20.00	.01

TABLE 2

MIDDLE-CLASS AND WORKING-CLASS IMPROVEMENT FROM NO-REWARD TO REWARD CONDITION

Class	*N*	Number of Items Correct			Right Minus Wrong		
		No Reward	Reward	*t*	No Reward	Reward	*t*
Middle	47	29.7	32.4	3.84***	23.8	25.4	1.83*
Working	27	26.2	29.5	5.35***	17.1	19.9	3.52***

* $p < .05$.
*** $p < .001$.

TABLE 3

PERCENTAGE OF MIDDLE-CLASS AND WORKING-CLASS *Ss* WITH LOW, MIDDLE, AND HIGH IMPROVEMENT

Class	Low[a]	Middle[a]	High[a]
Total correct			
Middle	46.8	21.3	31.9
Working	25.9	48.2	25.9
Right minus wrong			
Middle	46.8	19.2	34.0
Working	18.5	55.5	25.9

[a] Low, middle, and high cut-off points obtained by trisecting the total distribution based on all 150 *Ss* in the initial pool.

Results of the questionnaire on the desirability of the five-dollar reward indicate, as also reported by Douvan (1), that there was no class difference in conscious desire for the reward. Apparently the reward was large enough to elicit a high degree of motivation from the *Ss* regardless of class position. Therefore, the greater number of working-class *Ss* showing improvement can not be attributed simply to the fact that money is scarcer in their environment.

Achievement striving and test anxiety. A greater percentage of middle-class than working-class *Ss* obtained *lower* scores on the second testing when the reward was offered than on the first testing when there was no reward—25.5 per cent compared to 11.1 per cent with total-correct scoring

($p < .1$), and 38.3 per cent compared to 14.8 per cent with right-minus-wrong scoring ($p < .02$). These *S*s did more poorly in the reward condition even though (*a*) they were as motivated as the others to obtain the five dollar prize, as indicated by their questionnaire responses, (*b*) they had as much room for improvement, i.e., their initial scores were comparable to the initial scores of the others, and (*c*) they should have shown at least some improvement due to practice effect, since the two test forms were given only a week apart and contained items of highly similar content and level of difficulty. It therefore seems necessary to conclude, in line with recent test anxiety research, that the *S*s whose scores went down actually increased their strivings in response to the reward offer, but in so doing touched off efficiency-reducing emotional responses such as fears and expectations of failure.

Increased striving can also be inferred on the part of most of the remaining middle-class *S*s, since all but four of them made some improvement, especially since the middle-class improvers actually improved to a greater extent than the working-class improvers ($p < .1$ for total correct and $p < .2$ for right-minus-wrong scoring).

This analysis suggests that (*a*) most middle-class *S*s, contrary to the finding by Douvan mentioned earlier, increase their achievement strivings under external reward conditions and (*b*) the effect of such an increase in striving in the middle class is either to improve performance or to impair it. It would follow that the distribution of improvement scores for the middle class should be bimodal and highly variable. Both these expectations are borne out in our data. The middle-class distribution shows tendencies toward bimodality (Table 3) and it is more variable than the working-class distribution (*F* ratio significant at the .05 level with either total-correct or right-minus-wrong scoring).

McClelland (2) suggests that fantasy verbalization as well as test performance is depressed by anxiety about achievement and failure, and that as a result an *S* who is high on achievement striving may obtain spuriously low scores on his projective need Achievement measure if he is also high on anxiety. If such anxiety about achievement and failure can be aroused, or perhaps intensified, by an increase in striving induced by external reward, and if this occurs mainly in the middle class as the present study suggests, a number of middle-class *S*s should be expected to obtain lower need Achievement scores in reward than in non-reward conditions. This reason could explain Douvan's (1) finding that the mean need Achievement score for the middle class does not rise in the reward condition.

CONCLUSIONS

Both studies support the hypothesis that middle-class test performance is more highly generalized and less subject to improvement through increased striving for material reward than working-class performance. Study B, with before-after data, suggests further that although middle-class *S*s may have achievement strivings that are relatively highly internalized and thus generalized to situations which have no external reward, they are like the working-class *S*s in that these strivings *do* increase when a reward is offered. *The main difference between the two groups seems to be that while working-class striving and performance tend to rise uniformly in response to reward stimuli, in the middle-class reward-induced increases in striving may either raise the level of performance or touch off anxiety responses that lower it.* The source of this anxiety is a problem for further research. The literature on social class would suggest that, whether or not a prize were offered, the attempt to improve one's performance on a task designated as measuring the "ability to think" might readily engage the self-esteem needs of many middle-class *S*s; moreover, for some *S*s, attempting to improve may have this effect even with less ego-involving tasks. The resulting anxiety, which might be called "improvement anxiety" to distinguish it from other kinds of test anxiety, could have resulted in the performance decrements found here. Another possible explanation of these decrements is that the offer of a prize produced guilt feelings in those *S*s for whom competition with siblings and peers for adult rewards is synonomous with the expression of unacceptable hostile impulses. Such *S*s might have felt competitive despite the announcement by the adult that it was possible for everyone to win the prize. Doing poorly under these conditions could function as a guilt-avoiding mechanism.

REFERENCES

1. DOUVAN, E. Social status and success strivings. *J. abnorm. soc. Psychol.*, 1956, **52**, 219–224.
2. MCCLELLAND, D. C. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
3. ROSEN, B. C. The achievement syndrome. *Amer. soc. Rev.*, 1956, **21**, 203–211.

Received June 14, 1957.

How does our society react to the problems of mental illness? READ

SOCIAL CLASS AND MENTAL ILLNESS

A Community Study

By AUGUST B. HOLLINGSHEAD *and* FREDRICK C. REDLICH, *both of Yale University*. This is the first book to deal with the interrelationships between social class and mental illness in an urbanized community—in this case, New Haven, Conn. The authors have examined the social structure of the community, the psychiatric patients in treatment, the institutions where they are cared for, and the psychiatrists who treat them. They find the New Haven community characterized by a distinct class structure. Each class exhibits definite types of mental illness. Furthermore, each class reacts to the presence of mental illness in its members in different ways, and the treatment of psychiatric patients within the various classes differs accordingly.

Successive chapters tell the story of how members of the community become patients, how they and their families respond to psychiatric intervention, and how social class affects both therapists and patients. In the final chapters, the authors offer recommendations on what our society could do, if it wants to, about improving socially determined shortcomings of psychiatric practice.

1958. 442 pages. $7.50

Send for an examination copy

JOHN WILEY & SONS, Inc. 440 Fourth Avenue, New York 16, N. Y.

Important Evergreen Psychology Books

$1 and up

The same fine soft-cover editions used in college psychology courses throughout the country. Now at your bookstore.

Masochism in Modern Man by *Theodor Reik*. $1.95

A Study of Interpersonal Relations. New Contributions to Psychiatry, edited by *Patrick Mullahy*. $2.45

Psychoanalysis: Evolution and Development. A Review of Theory and Therapy, by *Clara Thompson* and *Patrick Mullahy*. $1.75

Black Anger. A unique psychoanalytic biography of a native African medicine man, by *Wulf Sachs*. $1.75

Education and Human Relations by *Ashley Montagu*. $1.45

The Forgotten Language. An Introduction to the Understanding of Dreams, Fairytales and Myths, by *Erich Fromm*. $1.75

Rebel Without a Cause. The hypnoanalysis of a criminal psychopath, by *Robert M. Lindner*. $1.75

The Biosocial Nature of Man by *Ashley Montagu*. $1.25

Listening with the Third Ear. The Inner Experience of a Psychoanalyst by *Theodor Reik*. $1.95

Democracy and Dictatorship. An incisive study of the psychological and social conditions that produce each, by *Zevedei Barbu*. $1.45

Jung's Psychology and Its Social Meaning by *Ira Progoff*. $1.75

Oedipus: Myth and Complex. A review of all major psychoanalytic theories by *Patrick Mullahy*. $1.95

Specialized Techniques in Psychotherapy. Edited by *Gustav Bychowski, M.D.* and *J. Louise Despert, M.D.* $1.95

The Marquis de Sade. An essay by *Simone de Beauvoir*, with selections from de Sade's writings. $1.75

SEND FOR FREE DESCRIPTIVE CATALOG

grove press
795 Broadway, New York 3

A NEW PUBLICATION

VOL. I, PART I JAN–MARCH 1958

LANGUAGE AND SPEECH

***Edited by* D. B. FRY**

This new quarterly journal is devoted principally to experimental research in the field of language and speech. The first issue is now available and contains the following papers:

KATHERINE S. HARRIS (*Haskins Laboratories, New York*). Cues for the discrimination of American English fricatives in spoken syllables.

G. HERDAN (*University of Bristol*). The relation between the functional burdening of phonemes and their frequency of occurrence.

A. R. LURIA (*University of Moscow*). Brain disorders and language analysis.

D. B. FRY AND P. DENES (*University College, London*). The solution of some fundamental problems in mechanical speech recognition.

FRIEDA GOLDMAN-EISLER (*University College, London*). Speech analysis and mental processes.

Annual Subscription: £1 (or $11.50).

Subscriptions should be sent to the publishers:
Robert Draper Ltd.
Kerbihan House, 85 Udney Park Road, Teddington, Middlesex, England.

Contributions should be addressed to:
D. B. Fry,
University College, Gower Street, London W.C.1

Neurology and Psychiatry

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles on neurological surgery in addition to general clinical neurology and experimental neurophysiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition the following special branches are dealt with fully: psychoanalysis, individual and analytical psychology; sexology, criminology, alcoholism, and drug addiction as related to psychiatry; psychosomatic medicine; social and industrial psychology and psychiatry, vocational guidance and mental testing; psychometrics; personality testing and the Rorschach test; heredity and statistical studies as they apply to psychiatry. The subjects of mental defect and epilepsy are fully reported.

Annual volume of 1000 pp., including an index classified by author and subject

Subscription price $22.50 per annual volume

The EXCERPTA MEDICA FOUNDATION

The New York Academy of Medicine Building

2 East 103 Street New York 29, N. Y.

STANFORD university press

Person Perception and Interpersonal Behavior

Edited by Renato Tagiuri and Luigi Petrullo. The first book in English to deal exclusively with these matters. The contributions by a distinguished body of psychologists, sociologists, and anthropologists systematically mark out areas of theory as well as of research. $7.50

Psychotherapy by Reciprocal Inhibition

Joseph Wolpe, M.D. The direct product of modern learning theory, this pioneer method of psychotherapy has attracted great attention because of the quick and lasting results it has achieved. $5.00

Psychology in the Soviet Union

Edited by Brian Simon. "The English editors had a difficult task in selecting, translating, and editing the Soviet contributions, and they have accomplished their task with excellence. . . . The articles are representative of a great diversity of psychological endeavor."—*Science.* $6.00

The Child within the Group

Marion E. Turner. AN EXPERIMENT IN SELF-GOVERNMENT. "It took a bit of a miracle to produce this book—in fact, a double miracle. That a teacher should have made this experience possible was unusual enough; that the teacher should also have recorded it is almost too good to be true."—*Theodore M. Newcomb,* in the Foreword. $3.00

Naven

Gregory Bateson. SECOND EDITION. This revised edition of an anthropological classic contains a new chapter which relates the basic ideas of the book to modern theories of psychiatry. $6.00

Stanford, California